

Introducción al tema de cointegración y tendencias

Alvaro Escribano Sáez

Universidad Complutense de Madrid

1. Introducción

El objetivo de esta panorámica es servir de complemento, y no de sustituto, de otras excelentes revisiones de la literatura que se han hecho últimamente. Véanse Dolado y Jenkinson (1987) y Stock y Watson (1987) y más recientemente Dolado (1989), Dolado *et. al.* (1989) y Engle y Yoo (1989).

En esta revisión se hace especial énfasis en el álgebra que hay que conocer para poder desarrollar y obtener los resultados fundamentales. De ahí el que esté escrito con un enfoque unificador y espero que didáctico, aunque su contenido no tiene porqué ser fácilmente asimilable ya que hay muchos conceptos poco conocidos para el no especialista.

La evolución histórica de la mayoría de las series macroeconómicas se caracteriza por su crecimiento. Las economías modernas están en constante expansión por diversas razones, entre las que cabría destacar el progreso técnico, los aumentos de población, etc. Las sendas temporales que siguen estas series distan mucho de ser constantes y lineales durante períodos largos de tiempo. Por el contrario siguen sendas oscilantes y cambiantes a lo largo de la tendencia. Ya los primeros analistas de series temporales pensaron que los procesos generadores de estas series podrían caracterizarse por tener diversos componentes como son el tendencial, el estacional, el cíclico y un elemento irregular. Hoy en día creo que se puede decir que hay un acuerdo unánime sobre que éste es un enfoque razonable. La discrepancia radica en la falta de unicidad en la descomposición de una serie en estos componentes no observables. Las diversas formas de parametrizar los componentes difieren en función de si los componentes son determinísticos, estocásticos o ambos a la vez, véase por ejemplo Maravall (1987).

El contenido de esta panorámica está organizado de la siguiente forma. En la sección 2 se indica cómo descomponer una serie univariante en sus términos transitorios (cíclicos) y permanentes (tendenciales). A su vez se estudia la relación que hay entre la existencia de raíces unitarias y las tendencias estocásticas.

En la sección 3, se estudian en un marco unificador la forma de reparametrizar los modelos de series temporales para aislar el término de la raíz unitaria. El análisis se extiende al caso de un número indeterminado de raíces unitarias.

En la sección 4, se presentan los resultados fundamentales de estimación e inferencia a nivel univariante cuando existen raíces unitarias. Estos resultados se utilizan para derivar las distribuciones de los contrastes de Dickey y Fuller y de Phillips.

La discusión general sobre el concepto de integración y sus relaciones con el concepto de tendencia se establece en la sección 5.

En la sección 6, se empieza el estudio multivariante. Se hace especial hincapié en las implicaciones de la cointegración. Se estudian las repercusiones, tanto en la representación de medias móviles como en la representación autorregresiva vectorial. Por último, se analizan las implicaciones de que las variables tengan tendencias en la media. Se introduce, por tanto, el concepto más general de co-tendencias en los momentos, y en especial el de co-tendencias en la media. La obtención de los modelos de corrección de error se realiza por dos vías alternativas, la primera se basa en reparametrización derivada de realizar una expansión de Taylor de primer orden en la matriz autorregresiva y la segunda se basa en la descomposición de Smith-McMillan-Yoo. De esta última representación se obtiene el concepto de cointegración polinomial.

Por último, en la sección 7, se estudian los problemas de estimación e inferencia en presencia de raíces unitarias y de variables cointegradas. En concreto se discuten los procedimientos de contrastación del número de raíces unitarias de los modelos multivariantes. En especial se mencionan los contrastes de Engle y Granger (1990), Stock y Watson (1988) y Johansen (1988).

2. Componentes permanentes y transitorios

En esta sección nos adentraremos en la discusión sobre la descomposición de una serie en componentes no observables y su relación con los modelos ARIMA. En concreto partiremos de modelos ARIMA (p, d, q) y añadiremos a la metodología sugerida por Box y Jenkins (1970), los contrastes sobre raíces unitarias que han aparecido en la literatura durante los últimos diez años.

La existencia de *raíces unitarias* se puede explicar, desde el punto de vista económico, como consecuencia de la existencia de *mercados eficientes*. Por ejemplo en los mercados financieros eficientes los cambios en precios deberían ser impredecibles.

Si x_t indica el precio de una acción en el período t , su representación podría ser

$$x_t = x_{t-1} + \varepsilon_t \quad (2.1)$$

donde ε_t es independiente e idénticamente distribuido con media cero y varianza σ^2 constante. ε_t es por tanto una innovación con respecto a la información que había en el mercado en el período $t - 1$.

Como la ecuación (2.1) es válida para todo t , también lo será que $t - 1$ y, por tanto, $x_{t-1} = x_{t-2} + \varepsilon_{t-1}$. Siguiendo con la sustitución recursiva hasta el período inicial, se obtiene de (2.1)

$$x_t = x_0 + \varepsilon_1 + \varepsilon_2 + \cdots \varepsilon_t \quad (2.2)$$

Nótese que en este caso cualquier innovación ε_i para $0 < i \leq t$, tendrá siempre impacto por muy distante que sea t . El camino aleatorio, ecuación (2.1),

no es más que una acumulación permanente de shocks aleatorios. Concluimos, por tanto, que la existencia de raíces unitarias implica que al menos hay shocks de tipo *permanente*.

Supongamos sin pérdida de generalidad, que $x_0 = \varepsilon_0 = 0$. La *media* de x_t es $E(x_t) = 0$. La *varianza de* x_t es $\text{var}(x_t) = E(\varepsilon_1 + \dots + \varepsilon_t)^2 = t\sigma^2$ ya que los términos $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$.

Consideremos ahora un camino aleatorio con deriva

$$x_t = b + x_{t-1} + \varepsilon_t \quad (2.3)$$

donde ε_t tiene las mismas propiedades del modelo (2.1).

Sustituyendo recursivamente en (2.3) obtenemos

$$x_t = x_0 + bt + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t \quad (2.4)$$

Nótese que, en este caso, los elementos permanentes no son sólo los estocásticos, ε_i , sino también los determinísticos, b . El término bt no es más que la acumulación permanente del término constante b durante t periodos. En este caso, la raíz unitaria genera un componente permanente estocástico y uno determinístico.

Bajo el supuesto de que $x_0 = \varepsilon_0 = 0$, la *media de* x_t es $E(x_t) = bt$ y la *varianza* $\text{var}(x_t) = E(x_t - bt)^2 = E(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t)^2 = t\sigma^2$.

La existencia de raíces unitarias implica que la serie tendrá una tendencia en la varianza, y a lo mejor también en la media, y por tanto, generará series que son *no estacionarias*. Sin embargo puede darse el caso de que la serie tenga una tendencia en alguno de los momentos pero no tenga una raíz unitaria, como ocurre en la ecuación (2.5). (Una discusión más rigurosa del concepto de tendencia se encuentra en Escribano (1987,a) y será discutida en la sección 5.)

Consideremos el siguiente modelo,

$$x_t = bt + \varepsilon_t \quad (2.5)$$

La media de x_t es $E(x_t) = bt$ y la varianza, $\text{var}(x_t) = E(x_t - bt)^2 = \sigma^2$, es constante. En este caso el elemento permanente es puramente determinístico, y el elemento estocástico ε_t es puramente *transitorio*, ya que sólo tiene un efecto contemporáneo sobre x_t pero no sobre los valores futuros de x_t .

Consideremos un caso más general donde $x_t \sim \text{ARIMA}(0, 1, \infty)$.

$$x_t = b + x_{t-1} + \theta(B)\varepsilon_t \quad (2.6)$$

donde $\theta(B)$ es un polinomio en el operador de retardos B . El operador de retardos es tal que $B^k x_t = x_{t-k}$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 + \dots$, y donde $\theta(B) = 0$ tiene todas las raíces fuera del círculo de la unidad, es decir, la media móvil es invertible.

Sustituyendo recursivamente en (2.6) obtenemos

$$x_t = x_0 + bt + \theta(B)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) \quad (2.7)$$

En forma de sumatorio se puede escribir como

$$x_t = x_0 + bt + \theta(B) \sum_{j=0}^{t-1} \varepsilon_{t-j} \quad (2.8)$$

El objetivo ahora es descomponer la serie x_t del modelo (2.8) en sus términos permanentes y transitorios. Para ello vamos a hacer uso de una expansión muy útil que es la expansión de Taylor. Como $\theta(B)$ es un polinomio en el operador de retardos B , las expansiones de Taylor siempre van a ser exactas cuando el término residual se evalúe en el punto B . Por lo tanto, las expansiones de Taylor nos van a dar formas de reparametrizar los polinomios $\theta(B)$.

Una reparametrización muy útil se obtiene tomando una expansión de Taylor de primer orden alrededor del punto $B = 1$ y evaluando el residuo en el punto B .

$$\theta(B) = \theta(1) + (1 - B)\theta^*(B) \quad (2.9)$$

donde $\theta^*(B) \equiv \frac{d\theta(B)}{dB} = 0$ tiene todas las raíces fuera del círculo de la unidad.

Sustituyendo (2.9) en (2.8) obtenemos

$$x_t = x_0 + bt + \theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + (1 - B)\theta^*(B) \sum_{j=0}^{t-1} \varepsilon_{t-j} \quad (2.10)$$

Como $x_0 = \varepsilon_0 = 0$, el término $(1 - B) \sum_{j=0}^{t-1} \varepsilon_{t-j} = \varepsilon_t$, de manera que la ecuación (2.10) se simplifica de la forma siguiente:

$$x_t = bt + \theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + \theta^*(B)\varepsilon_t \quad (2.11)$$

Como $\theta^*(B) = 0$ tiene todas las raíces fuera del círculo de la unidad, los términos ε_t ya no tendrán un efecto permanente ya que su impacto irá decayendo a medida que nos alejemos en el tiempo. Por tanto, $\theta^*(B)\varepsilon_t$ representa el *componente cíclico* (el denominado ciclo de actividad económica) y sus efectos serán transitorios.

Sin embargo, los términos bt y $\theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j}$ representan los elementos permanentes, ya que los efectos determinísticos de b y los estocásticos de $\theta(1)\varepsilon_t$, siempre tendrán impacto. A los elementos permanentes los vamos a incluir en el *componente tendencial* τ_t , de forma que

$$x_t = \tau_t + \theta^*(B)\varepsilon_t \quad (2.12)$$

donde $\tau_t = bt + \theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j}$.

Nótese que τ_t tiene una representación similar a la ecuación (2.4), que fue obtenida de un camino aleatorio con deriva, donde $\tau_t = b + \tau_{t-1} + \theta(1)\varepsilon_t$. Utilizando esta expresión en (2.12) obtenemos la conocida descomposición de Beveridge y Nelson (1981),

$$x_t = \tau_t + \theta^*(B)\varepsilon_t \quad (2.13a)$$

$$\tau_t = b + \tau_{t-1} + \eta_t \quad (2.13b)$$

$$\eta_t = \theta(1)\varepsilon_t \quad (2.13c)$$

El componente tendencial τ_t tiene elementos determinísticos, b , y elementos estocásticos, η_t , y por eso se le denomina *tendencia estocástica*, Watson (1986). La media de esta tendencia estocástica es $E(\tau_t) = bt$ y la varianza $\text{var}(\tau_t) = E(\tau_t - bt)^2 = t\theta(1)^2\sigma^2$. Nótese que en este caso los shocks transitorios ε_t están relacionados con los shocks permanentes η_t a través de la ecuación (2.13c).

La descomposición de una serie macroeconómica en sus componentes permanente y transitorio ha tenido una gran importancia en economía desde la contribución de Friedman (1957) con la teoría de la renta permanente. En la actualidad el tema de interés trata de identificar si los shocks permanentes y los transitorios están relacionados entre sí o por el contrario son independientes, ver Stock y Watson (1987). Como hemos visto antes, la existencia de raíces unitarias genera tendencias estocásticas y shocks permanentes. En la siguiente sección discutiremos cómo se puede identificar el número de raíces unitarias que hay en las series.

3. Contrastes de identificación del número de raíces unitarias a nivel univariante

En esta sección presentaremos algunos contrastes de raíces unitarias que deberían añadirse a la metodología de Box-Jenkins (1976), para ayudar a determinar el número de veces que hay que diferenciar una serie antes de pasar a identificar el orden del modelo ARMA. Como veremos más adelante estos contrastes no están exentos de problemas y por eso deben utilizarse como complemento de la metodología antes citada y no como sustituto.

Partimos de que x_t satisface el teorema de representación de Wold,

$$(1 - B)^d(x_t - \mu_t) = \theta(B)\varepsilon_t \quad (3.1)$$

donde $\theta(B)$ es un polinomio en el operador de retardos B . ε_t es ruido blanco con media cero y varianza σ^2 constante. La ecuación $\theta(B) = 0$ tiene todas las raíces fuera del círculo de la unidad y μ_t es la media de x_t . Sin pérdida de generalidad vamos a considerar el caso $d = 1$ ya que es el más común cuando se estudian series macroeconómicas. Posteriormente mencionaremos cómo se extiende el análisis al caso general.

Cuando $d = 1$ la ecuación (3.1) se reduce a

$$(1 - B)(x_t - \mu_t) = \theta(B)\varepsilon_t \quad (3.2)$$

Si consideramos que la media de x_t es bt , la ecuación (3.2) no es más que una reparametrización de la ecuación (2.6) del apartado anterior. Para llevar a cabo los contrastes de raíces unitarias es conveniente escribir los modelos en forma autorregresiva. Para ello multiplicaremos a (3.2) por $\theta(B)^{-1} = a^*(B)$

$$a^*(B)(1 - B)(x_t - \mu_t) = \varepsilon_t \quad (3.3)$$

donde $a^*(B) = 0$ tiene todas las raíces fuera del círculo de la unidad.

La representación autorregresiva (3.3) ya está diferenciada una vez, $d = 1$. En la práctica no sabemos el número de veces que hay que diferenciar las series y, por tanto, vamos a partir de una *representación autorregresiva* sin restringir del tipo

$$a(B)(x_t - \mu_t) = \varepsilon_t \quad (3.4)$$

donde $a(B) = 0$ puede tener algunas raíces en el círculo de la unidad (raíces unitarias).

Para obtener los contrastes de Dickey-Fuller aplicamos una *expansión de Taylor* de primer orden alrededor del punto $B = 1$ al polinomio $a(B)$,

$$a(B) = a(1) + (1 - B)a^+(B) \quad (3.5)$$

donde $a^+(B) \equiv \frac{da(B)}{dB}$. Para $d = 1$, $a^+(B) = 0$ tendrá todas las raíces fuera del círculo de la unidad. Sumando y restando $a(1)B$ en (3.5) se obtiene

$$a(B) = a(1)B + (1 - B)[a(1) + a^+(B)] = a(1)B + (1 - B)a^*(B) \quad (3.6)$$

Sustituyendo (3.6) en (3.4), obtenemos

$$a^*(B)(1 - B)(x_t - \mu_t) = -a(1)(x_{t-1} - \mu_{t-1}) + \varepsilon_t \quad (3.7)$$

Si $d = 1$, de forma que hay una raíz unitaria y por tanto hay que diferenciar una vez, entonces $a(1) = 0$ y (3.7) se reducirá a (3.3). Si, en cambio, $a(1) > 0$ al diferenciar una vez estaremos sobrediferenciando a no ser que incorporemos el término $(x_{t-1} - \mu_{t-1})$ en la ecuación.

Intuitivamente la idea del contraste de raíces unitarias se puede explicar con un ejemplo.

Sea x_t un AR(1) con media $\mu_t = 0$

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.8)$$

Restando x_{t-1} a ambos lados de la ecuación

$$(1 - B)x_t = (\rho - 1)x_{t-1} + \varepsilon_t \quad (3.9)$$

o lo que es lo mismo,

$$(1 - B)x_t = -(1 - \rho)x_{t-1} + \varepsilon_t \quad (3.10)$$

Comparando (3.10) con (3.7) observamos que $(1 - \rho) = a(1)$. $a(1)$ será igual a cero si $\rho = 1$ y, por tanto, habrá una raíz unitaria en el AR (1) de la ecuación (3.8). Si por el contrario $a(1) > 0$, entonces $\rho < 1$ y el modelo (3.8) será estable. En términos de la sección anterior si $\rho = 1$ los shocks, ε_t , serán de tipo permanente, mientras que si $\rho < 1$ los shocks, ε_t , serán transitorios.

En la práctica la media de x_t , μ_t , es desconocida y tendremos que estimarla. Reescribiendo (3.7) sin desviaciones con respecto a la media

$$a^*(B)(1 - B)x_t = a^*(B)(1 - B)\mu_t + a(1)\mu_{t-1} - a(1)x_{t-1} + \varepsilon_t \quad (3.11)$$

Bajo la hipótesis nula de que hay una raíz unitaria, $H_0: a(1) = 0$, la ecuación (3.11) se reduce a

$$a^*(B)(1 - B)x_t = a^*(B)(1 - B)\mu_t + \varepsilon_t \quad (3.12)$$

Esta ecuación es una representación autorregresiva donde todos los componentes, menos el término del ruido blanco ε_t , están en primeras diferencias.

La ecuación (3.11) es la representación en base a la cual se desarrolla el contraste de *Dickey y Fuller*. Según se consideren distintas medias μ_t , el contraste adoptará una forma u otra como se verá más adelante. Para obtener este contraste de forma explícita vamos a desarrollar el polinomio $a^*(B)$, partiendo de que $a^*(0) = 1$ y por tanto la ecuación (3.11) está en forma reducida,

$$a^*(B) = 1 - a^{**}(B)B \quad (3.13)$$

Sustituyendo (3.13) en (3.11) y reordenando términos, obtenemos

$$\begin{aligned} (1 - B)x_t &= (1 - B)\mu_t + a(1)\mu_{t-1} - a(1)x_{t-1} + a^{**}(B)(1 - B)\mu_{t-1} + \\ &+ a^{**}(B)(1 - B)x_{t-1} + \varepsilon_t \end{aligned} \quad (3.14)$$

Dickey y Fuller (1979) proponen contrastar la hipótesis nula $H_0: a(1) = 0$ mediante el *t*-ratio correspondiente al parámetro de la variable x_{t-1} en la regresión (3.11). Como la hipótesis nula es de hecho una hipótesis conjunta, de los coeficientes de μ_{t-1} y de x_{t-1} de la regresión (3.11), Dickey y Fuller (1981) proponen realizar el *contraste conjunto* basado en el contraste- \mathcal{F} . La distribución asintótica de estos estadísticos no es la estándar y será discutida en la siguiente sección.

Otro estadístico que ha tenido mucho impacto es el propuesto por Phillips (1987). Para obtener este contraste de forma explícita vamos a reescribir la ecuación (3.14) de la siguiente forma

$$(1 - B)x_t = (1 - B)\mu_t + a(1)\mu_{t-1} - a(1)x_{t-1} + u_t \quad (3.15)$$

donde $u_t \equiv a^{**}(B)(1 - B)\mu_{t-1} + a^{**}(B)(1 - B)x_{t-1} + \varepsilon_t$. Phillips (1987) y Phillips y Perron (1988) proponen estimar la hipótesis nula $H_0: a(1) = 0$ en base a modificaciones no paramétricas de los contrastes de Dickey-Fuller (1979, 1981), mediante la estimación de la regresión (3.15) como veremos en la siguiente sección.

Ahora nos planteamos el cómo obtener términos de medias móviles en los contrastes de raíces unitarias. Para ello aproximamos la representación autorregresiva, ecuación (3.4), por un ARMA (p, q)

$$a(B)(x_t - \mu_t) = \theta(B)\varepsilon_t \quad (3.16)$$

donde $a(B) = 1 - \sum_{j=1}^p a_j B^j$ y $\theta(B) = 1 - \sum_{j=1}^q \theta_j B^j$. $a(B) = 0$ puede tener algunas raíces en el círculo de la unidad pero no $\theta(B) = 0$. Supongamos que $\theta(B)$ es invertible y que no tiene factores comunes con $a(B)$.

Siguiendo los pasos de las ecuaciones (3.5) y (3.6), podemos escribir (3.16) como

$$a^*(B)(1 - B)x_t = a^*(B)(1 - B)\mu_t + a(1)\mu_{t-1} - a(1)x_{t-1} + \theta(B)\varepsilon_t \quad (3.17)$$

donde aparecen los términos de medias móviles explícitamente.

Bajo la hipótesis nula $H_0: a(1) = 0$ la representación (3.17) se reduce a

$$a^*(B)(1 - B)x_t = a^*(B)(1 - B)\mu_t + \theta(B)\varepsilon_t \quad (3.18)$$

que es un ARMA (p, q) en primeras diferencias.

La representación (3.18) nos va a servir para ejemplificar uno de los mayores problemas de los contrastes de raíces unitarias; su falta de poder.

Supongamos que $\theta(B) = 1 - \theta_1 B$, esto es $q = 1$ y por lo tanto, la MA (q) se ha convertido en una MA (1),

$$a^*(B)(1 - B)x_t = a^*(B)(1 - B)\mu_t + (1 - \theta_1 B)\varepsilon_t \quad (3.19)$$

Si θ_1 es próximo a la unidad, el factor $(1 - \theta_1 B)$ es próximo a $(1 - B)$ y, por tanto, el modelo (3.16) se va a parecer mucho en la práctica a un modelo sin raíces unitarias del tipo

$$a^*(B)x_t = a^*(B)\mu_t + \varepsilon_t \quad (3.20)$$

Este tipo de problemas fueron analizados en Molinas (1986) y en Schwert (1987) por medio de simulaciones.

Como es bien sabido toda media móvil se puede aproximar por un autorregresivo. Esta idea llevó a Said y Dickey (1984) a sugerir que el orden del proceso autorregresivo $a^*(B)$ en (3.8) aumente con el tamaño muestral (T), de forma que el término $\theta(B)\varepsilon_t$ de (3.17) se pueda aproximar por términos autorregresivos de orden p , donde el orden p es como máximo igual a $\sqrt[3]{T}$.

Cuando p y q , los órdenes de la parte autorregresiva y de la media móvil, son o bien conocidos a priori, o bien identificados por medio de las técnicas de Box-Jenkins (1976), Said y Dickey (1985) proponen contrastar $H_0: a(1) = 0$ directamente en la representación (3.17).

Hall (1989) propone contrastar la hipótesis nula en (3.17) pero estimando los parámetros por medio de variables instrumentales. Dolado (1989) discute la elección óptima de las variables instrumentales a utilizar en el caso de un ARMA (1, 1). En la sección 2 presentamos un ejemplo en el que la serie x_t tenía una tendencia en la media pero no una raíz unitaria, ecuación (2.5). Si en ese caso diferenciamos la serie para quitar la tendencia estaremos generando una raíz unitaria en la media móvil,

$$(1 - B)x_t = b + (1 - B)\varepsilon_t \quad (3.21)$$

Phillips y Ouliaris (1988) proponen un contraste de raíces unitarias basado en la parte de medias móviles en vez de en el componente autorregresivo como hemos discutido hasta ahora. El contraste se basa en comparar las varianzas de largo plazo de x_t , con la de $(1 - B)x_t$. En el caso de la ecuación (2.5) vimos que la varianza de x_t es σ^2 , mientras que la varianza de $(1 - B)x_t$ en (3.18) es igual a la varianza de $(1 - B)\varepsilon_t$ e igual a cero. Si x_t tiene una raíz unitaria $(1 - B)x_t$ tendrá una varianza de largo plazo distinta de cero; mientras que si x_t tiene todas sus raíces menores que la unidad en valor absoluto, entonces $(1 - B)x_t$ tendrá varianza cero. El contraste, por tanto, está basado en la sobrediferenciación de la serie, mediante la comparación de las varianzas.

Las implicaciones de infradiferenciar o sobrediferenciar una serie son muy distintas, véase por ejemplo Peña (1990) en este mismo número de la revista. Infradiferenciar equivale a identificar un número de raíces unitarias menor al que hay en el proceso generado de datos. Sobrediferenciar significa identificar un número de raíces unitarias mayor que el número de raíces unitarias del proceso generador. A continuación vamos a discutir los contrastes de identificación del número de raíces unitarias que hay en un proceso estocástico.

Supongamos que x_t está generado por un proceso ARMA (p, q) como el de la ecuación (3.16). En forma explícita podemos escribirlo como:

$$(x_t - \mu_t) = \sum_{j=1}^p a_j(x_{t-j} - \mu_{t-j}) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (3.22)$$

La ecuación característica de la parte autorregresiva es $m^p - \sum_{j=1}^p a_j m^{p-j} = 0$ cuyas raíces son m_1, m_2, \dots, m_p . La parte autorregresiva tiene algunas raíces unitarias de forma que $1 \geq |m_1| \geq |m_2| \dots \geq |m_p|$. La media móvil es invertible de forma que $1 > |m_1^*| \geq \dots \geq |m_q^*|$ donde $m_1^*, m_2^*, \dots, m_q^*$ son las raíces de la ecuación característica de la media móvil.

El polinomio de la parte autorregresiva $a(B) \equiv 1 - \sum_{j=1}^p a_j B^j$ se puede descomponer como el producto

$$a(B) = \prod_{i=1}^p (1 - m_i B) \quad (3.23)$$

Pantula (1989) probó el siguiente resultado,

Lema 3.1: Sea $(x_t - \mu_t)$ generado por el proceso (3.22). Entonces $m_1 = m_2 = \dots m_d = 1$, $d \leq p$, si y sólo si $\beta_1 = \beta_2 = \dots = \beta_d = 0$ en el siguiente modelo:

$$(1 - B)^p(x_t - \mu_t) = - \sum_{j=1}^p \beta_j(1 - B)^{j-1}(x_{t-1} - \mu_{t-1}) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (3.24)$$

Este lema relaciona los coeficientes β_j de la regresión (3.24) con las raíces del polinomio autorregresivo (3.23).

Como los componentes de la media móvil se pueden aproximar por términos autorregresivos que aumentan en función del tamaño muestral, Pantula (1989) propone la siguiente implementación del contraste de la hipótesis nula $H_0: \beta_1 = \beta_2 = \dots = \beta_d = 0$, frente a la alternativa $H_1: \beta_1 < 0, \beta_2 < 0, \dots, \beta_d < 0$.

$$(1 - B)^r(x_t - \mu_t) = - \sum_{j=1}^r \beta_j(1 - B)^{j-1}(x_{t-1} - \mu_{t-1}) + \sum_{j=1}^{k-r} b_j(1 - B)^r(x_{t-j} - \mu_{t-j}) + \varepsilon_t \quad (3.25)$$

donde $r = 1, 2, \dots, d$ y $k \geq p$ donde k es como máximo $\sqrt[3]{T}$. La hipótesis conjunta se contrasta mediante el estadístico \mathcal{F} para valores de r que van de mayor a menor.

Cuando $r = 1$, el estadístico \mathcal{F} es el cuadrado del t -ratio considerado por Said y Dickey (1984).

Un contraste de mayor poder que el estadístico \mathcal{F} es el basado en el t -ratio. La hipótesis nula $H_0: \beta_1 = \beta_2 = \dots = \beta_d = 0$ se contrasta frente a la alternativa $H_1: \beta_1 = \beta_2 = \dots = \beta_{d-1} = 0$ y $\beta_d < 0$, en

$$(1 - B)^d(x_t - \mu_t) = -\beta_d^*(1 - B)^{d-1}(x_{t-1} - \mu_{t-1}) + \sum_{j=1}^{k-d} b_j^*(1 - B)^d(x_{t-j} - \mu_{t-j}) + \varepsilon_t \quad (3.26)$$

Bajo $H_0: \beta_d^* = 0$ y bajo $H_1: \beta_d^* < 0$. $k \geq p$ y k es como máximo $\sqrt[3]{T}$.

Para $d = 1$ este estadístico basado en el t -ratio es el propuesto por Said y Dickey (1984).

Si el proceso generador de datos es un autorregresivo puro entonces $k = p$ y los contrastes se basan también en las ecuaciones (3.25) y (3.26), respectivamente.

Los otros procedimientos para contrastar raíces unitarias en presencia de medias móviles, que vimos antes para el caso $d = 1$ se podrían generalizar de forma similar para d raíces unitarias. Al igual que hicimos para el caso $d = 1$ podemos expresar la ecuación (3.26) sin desviaciones con respecto a la media,

$$(1 - B)^d(x_t - \mu_t) = (1 - B)^d\mu_t + \beta_k^*(1 - B)^{d-1}\mu_{t-1} - \beta_k^*(1 - B)^{d-1}x_{t-1} +$$

$$- \sum_{j=1}^{k-d} b_j^*(1-B)^d \mu_{t-j} + \sum_{j=1}^{k-d} b_j^*(1-B)^d x_{t-j} + \varepsilon_t \quad (3.27)$$

Nótese que bajo la hipótesis conjunta $H_0: \beta_k^* = 0$, esta ecuación se reduce a un proceso autorregresivo en variables diferenciadas d veces,

$$(1-B)^d x_t = (1-B)^d \mu_t - \sum_{j=1}^{k-d} b_j^*(1-B)^d \mu_{t-j} + \sum_{j=1}^{k-d} b_j^*(1-B)^d x_{t-j} + \varepsilon_t \quad (3.28)$$

Otra ventaja de contrastar las raíces unitarias en base a las ecuaciones (3.26) y (3.27), es que los valores críticos de los t -ratios y del estadístico \mathcal{F} , de (3.27), son los mismos que los generados mediante simulaciones de Monte-Carlo por Dickey y Fuller (1979 y 1981), ya que las distribuciones de los estadísticos basados en las ecuaciones (3.24) a (3.27) no siguen distribuciones estándar.

En la implementación de estos contrastes a observaciones provenientes de variables económicas, es importante tener en cuenta que la media μ_t es distinta de cero. En períodos largos de tiempo es incluso probable que la media μ_t varíe bruscamente debido a cambios estructurales. Las implicaciones de esos cambios estructurales sobre los contrastes de raíces unitarias han sido tratados en Andrés *et. al.* (1990), Escribano (1990) y Perron (1989).

4. Estimación e inferencia en la identificación de las raíces unitarias

Las propiedades de series temporales univariantes tienen importantes implicaciones en la estimación de parámetros así como en la contrastación de hipótesis. Este hecho ha servido para unir dos enfoques que durante los años 70 estaban separados. Por un lado el análisis de series temporales y por el otro los modelos de regresión considerados por los económetras. En este apartado vamos a ver cómo la estimación e inferencia en modelos univariantes depende en gran medida de los componentes no observables que dominan la serie. En la sección sexta nos dedicaremos a los modelos de regresión.

Partimos de un modelo como el de la ecuación (3.15) con $a(1) = 0$,

$$(1-B)(x_t - \mu_t) = u_t \quad (4.1)$$

donde u_t puede tener diferentes representaciones. Si $u_t = a(\beta)^{-1}\theta(B)\varepsilon_t$, tendremos que $(x_t - \mu_t)$ es un ARIMA $(p, 1, q)$. Si u_t es independiente e idénticamente distribuido con media cero y varianza constante σ^2 , entonces $(x_t - \mu_t)$ es un camino aleatorio. u_t puede ser un modelo del tipo $u_t = a(B)^{-1}\theta(B)\varepsilon_t + a(B)^{-1}b(B)d(B)^{-1}f(B)v_t$. En este caso, la representación de $(x_t - \mu_t)$ es un ARIMA $(p, 1, q)$ con variables exógenas, z_t , ver Perron (1986) y Andrews (1988),

$$a(B)(1-B)(x_t - \mu_t) = b(B)z_t + \theta(\beta)\varepsilon_t \quad (4.2.a)$$

$$d(B)z_t = f(B)v_t \quad (4.2.b)$$

Queremos ver qué tipo de condiciones generales sobre u_t incluyen toda esta posible gama de modelos y a la vez nos permiten estimar $a(1)$ en una ecuación como la siguiente

$$(1 - B)(x_t - \mu_t) = -a(1)(x_{t-1} - \mu_{t-1}) + u_t \quad (4.3)$$

sin necesidad de hacer explícito el modelo que está generando el término estocástico u_t . Las condiciones más generales fueron las propuestas por Phillips (1987) y son las que vamos a discutir en esta sección.

Supuesto 1

El proceso estocástico $x_t - \mu_t$, generado por (4.1) satisface las siguientes condiciones:

- a) $E(u_t) = 0$ para todo t .
- b) $\sup_t E|u_t|^\beta < \infty$ para algún $\beta > 2$.
- c) $\sigma^2 = \lim_{T \rightarrow \infty} E(T^{-1}S_T^2)$ existe y $\sigma^2 > 0$ donde $S_T = \sum_{j=0}^{T-1} u_{T-j}$
- d) u_t es una α -mezcla (mezcla en sentido fuerte) con coeficientes de mezcla α_z que satisfacen $\sum_{\tau=0}^{\infty} \alpha_\tau^{(1-2/\beta)} < \infty$.

A continuación vamos a estudiar las implicaciones de cada uno de estos supuestos. Para ello vamos a reescribir la ecuación (4.1) en la forma de la ecuación (2.2) mediante sustitución recursiva

$$(x_t - \mu_t) = (x_0 - \mu_0) + u_1 + u_2 + \dots + u_t \quad (4.4)$$

Supondremos sin pérdida de generalidad, que $(x_0 - \mu_0) = 0$ y denominaremos $S_t = u_1 + u_2 + \dots + u_t$, de forma que (4.4) se reduce a

$$(x_t - \mu_t) = S_t \quad (4.5)$$

La condición a) no es restrictiva ya que el modelo está en desviaciones con respecto a la media. La condición b) garantiza que el momento de orden $\beta > 2$ existe. Por Sup se entiende el supremo, que es la mínima cota superior. Se utiliza el supremo en vez del máximo por razones de generalidad. Como u_t no necesita ser un proceso estacionario, esta condición controla el grado de heterogeneidad permitido. La condición c) es una condición de convergencia en el promedio de la varianza de la suma parcial, S_t . Para explicar lo que indica la condición d) necesitamos primero definir lo que se entiende por α -mezcla.

Sea $\mathcal{B}_{-\infty}^t \equiv \sigma(\dots, x_t)$ el conjunto de información (σ -álgebra) que contiene el pasado y el presente de x_t , y sea $\mathcal{B}_{t+\tau}^\infty \equiv \sigma(x_{t+\tau}, \dots)$ el conjunto de información que contiene el futuro de la serie x_t desde el instante $t + \tau$ en adelante. Si A y B

son dos eventos independientes, entonces $P(A \cap B) = P(A)P(B)$. Si no son independientes el grado de dependencia en probabilidad vendrá medido por la diferencia $P(A \cap B) - P(A)P(B)$.

Sea $\alpha(\mathcal{B}_{-\infty}^+, \mathcal{B}_{i+\tau}^\infty) \equiv \sup_{[A \in \mathcal{B}_{-\infty}^+, B \in \mathcal{B}_{i+\tau}^\infty]} |P(A \cap B) - P(A)P(B)|$. Si los conjuntos de

información $\mathcal{B}_{-\infty}^+$ y $\mathcal{B}_{i+\tau}^\infty$ no tienen nada en común, entonces $\alpha(\mathcal{B}_{-\infty}^+, \mathcal{B}_{i+\tau}^\infty) = 0$. Al coeficiente $\alpha(\tau) \equiv \sup_i \alpha(\mathcal{B}_{-\infty}^+, \mathcal{B}_{i+\tau}^\infty)$ se le denomina el *coeficiente de α -mezcla*, o de mezcla fuerte. Si $\alpha(\tau) \rightarrow 0$ cuando $\tau \rightarrow \infty$, entonces diremos que x_t es una α -mezcla, ver Rosenblatt (1978). Vemos por tanto que el concepto de α -mezcla mide el grado de dependencia. La condición d) del supuesto 1, controla el grado de dependencia temporal permitido en la serie u_t . Esta condición se satisface si $\alpha(\tau) = O(\tau^{-\lambda})$ donde $\lambda > \left(\frac{\beta}{\beta-2}\right)$. Cuanto más cerca de 2 esté β menor será la dependencia permitida, o lo que es lo mismo cuanto mayor sea el orden del momento de u_t que exista mayor será la posible dependencia temporal de u_t . Véase White (1984) para una excelente discusión de este concepto.

Nos vamos a concentrar en el espacio $D(0, 1)$ de todas las funciones valoradas en las reales, en el intervalo $[0, 1]$, y tal que sean continuas por la derecha y tengan límites finitos por la izquierda. Luego vemos, que se permiten algunas discontinuidades.

Lo primero, por tanto, es reducir los elementos de la ecuación (4.5) al intervalo $[0, 1]$, y esto se consigue mediante la siguiente estandarización.

$$X_T(r) = \frac{T^{-1/2}}{\sigma_x} S_{j-1} \quad \frac{j-1}{T} \leq r < \frac{j}{T} \quad \text{donde } j = 1, 2, \dots, T \quad (4.6)$$

$$X_T(1) = \frac{T^{-1/2}}{\sigma_x} S_T$$

Ahora $X_T(r)$ ya es un elemento del espacio funcional $D(0, 1)$.

Lema 4.1. (Teorema Central del Límite sobre funcionales, debido a Herrndorf.) Si u_t satisface el supuesto 1 entonces cuando $T \rightarrow \infty$, $X_T(r)$ converge en distribución a $B(r)$, un movimiento Browniano, también denominado proceso de Wiener.

Lema 4.2. (Teorema de la Proyección Continua, ver Billingsley (1986).) Si $X_T(r)$ converge en distribución a $B(r)$ cuando $T \rightarrow \infty$ y si g es una función continua en $D(0, 1)$, entonces $g[X_T(r)]$ converge en distribución a $g[B(r)]$ cuando $T \rightarrow \infty$.

Nuestro objetivo es conocer la distribución del t -ratio del parámetro $a(1)$ de la ecuación (4.3). Para ello primero estimamos el parámetro $a(1)$ en esa ecuación por mínimos cuadrados ordinarios y obtenemos

$$\hat{a}(1) = \frac{\frac{1}{T} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})(1 - B)(x_t - \mu_t)}{\frac{1}{T} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2} \quad (4.7)$$

El error estándar de $\hat{a}(1)$ estimado es

$$\hat{\sigma}_{\hat{a}(1)} = S \left(\sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2 \right)^{-1/2} \quad (4.8)$$

donde

$$S^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2 \quad (4.9)$$

por lo tanto, el t -ratio de $\hat{a}(1)$ será

$$t_{\hat{a}(1)} = \frac{\hat{a}(1) - 0}{\hat{\sigma}_{\hat{a}(1)}}$$

Phillips (1987) probó los siguientes resultados que nos serán útiles para derivar la distribución asintótica de $\hat{a}(1)$ y de $t_{\hat{a}(1)}$.

Lema 4.3. Si u_t satisface el supuesto 1 y si $\text{Sup}_t E|u_t|^{\beta+\eta} < \infty$ para algún $\eta > 0$ y $\beta > 2$, entonces cuando $T \rightarrow \infty$.

a) $T^{-2} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2$ converge en distribución a $\sigma^2 \int_0^1 B(r)^2 dr$.

b) $T^{-1} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})(1 - B)(x_t - \mu_t)$ converge en distribución a

$$\left(\frac{\sigma^2}{2} \right) \left(B(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right)$$

c) S^2 converge en probabilidad a σ_u^2 .

Para obtener la distribución de $\hat{a}(1)$ multiplicamos a la ecuación (4.7) por T .

$$T\hat{a}(1) = \frac{\frac{1}{T} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})(1 - B)(x_t - \mu_t)}{\frac{1}{T^2} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2} \quad (4.10)$$

Aplicando el lema 4.2 y el lema 4.3 a) y b) obtenemos de forma inmediata que

$$T\hat{a}(1) \text{ converge en distribución a } \frac{\left(\frac{1}{2}\right)\left(B(1)^2 - \frac{\sigma_u^2}{\sigma^2}\right)}{\int_0^1 B(r)^2 dr} \quad (4.11)$$

De igual forma aplicando el lema 4.2 y el lema 4.3 a), b) y c), obtenemos que

$$t_{\hat{a}(1)} \text{ converge en distribución a } \frac{\left(\frac{\sigma}{2\sigma_u}\right)\left(B(1)^2 - \frac{\sigma_u^2}{\sigma^2}\right)}{\left[\int_0^1 B(r)^2 dr\right]^{1/2}} \quad (4.12)$$

De la expresión (4.11) concluimos que $\hat{a}(1)$ converge en probabilidad a cero a la tasa $Op(T^{-1})$ en vez de a la convencional $Op(T^{-1/2})$. Al ser $\hat{a}(1)$ convergente y a una velocidad mayor que la estándar se le denomina «*superconsistente*».

A su vez en la expresión (4.12) se ve claramente por qué el t -ratio de $\hat{a}(1)$ no sigue una distribución normal asintóticamente. Es una *distribución no estándar* que depende de parámetros que no son de interés, como por ejemplo σ_u^2 y σ^2 donde

$$\sigma_u^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(u_t)^2 \quad (4.13)$$

y

$$\sigma_T^2 = \text{var}(T^{-1/2}S_T) = \text{var}[T^{1/2}(x_t - \mu_t)] = E[T^{-1}S_T^2]$$

Desarrollando el último término obtenemos

$$\sigma_T^2 = T^{-1} \sum_{t=1}^T E(u_t^2) + T^{-1} 2 \sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^T E(u_t u_{t-\tau}) \quad (4.14)$$

Cuando la contribución de $E(u_t u_{t-\tau})$ es pequeña para τ grande, White (1984) propone aproximar σ_T^2 por σ_{Tl}^2 donde l es el retardo de truncamiento

$$\sigma_{Tl}^2 = T^{-1} \sum_{t=1}^T E(u_t^2) + T^{-1} 2 \sum_{\tau=1}^l \sum_{t=\tau+1}^T E(u_t u_{t-\tau}) \quad (4.15)$$

Bajo la condición de que el $\text{Sup}_t E|u_t|^{2\beta} < \infty$ para $\beta > 2$ y si $l = O(T^{1/4})$ el estimador

$$S_{Tl}^2 = T^{-1} \sum_{t=1}^T u_t^2 + T^{-1} 2 \sum_{\tau=1}^l \sum_{t=\tau+1}^T u_t u_{t-\tau} \quad (4.16)$$

converge en probabilidad a σ^2 . El estimador S_{Tl}^2 no garantiza que el resultado sea no negativo. Para resolver este problema Newey y West (1987) proponen introducir las ponderaciones $w_{t\tau}$ en el estimador S_{Tl}^2 .

$$\tilde{S}_{Tl}^2 = T^{-1} \sum_{t=1}^T u_t^2 + T^{-1} 2 \sum_{\tau=1}^l w_{t\tau} \sum_{t=\tau+1}^T u_t u_{t-\tau} \quad (4.17)$$

donde $w_{t\tau} = 1 - \frac{\tau}{l+1}$.

En vista de que los parámetros que no son de interés, ver ecuaciones (4.11) y (4.12), se pueden estimar consistentemente con las ecuaciones (4.13) y (4.17), Phillips (1987) propone corregir los estadísticos $T\hat{a}(1)$ y $t_{\hat{a}(1)}$ de forma que su distribución límite no dependa ni de σ^2 ni de σ_u^2 .

$$Z_{\hat{a}(1)} = T\hat{a}(1) - \frac{\left(\frac{1}{2}\right)(\tilde{S}_{Tl}^2 - S_u^2)}{T^{-2} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2} \quad (4.18)$$

donde $S_u^2 = T^{-1} \sum_{t=1}^T [(1-B)(x_t - \mu_t)]^2 = T^{-1} \sum_{t=1}^T u_t^2$.

$$Z_{t_{\hat{a}(1)}} = \frac{\hat{a}(1) - 0}{S_{Tl} \left(\sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2 \right)} - \frac{\left(\frac{1}{2}\right)(\tilde{S}_{Tl}^2 - S_u^2)}{S_{Tl} \left(T^{-2} \sum_{t=1}^T (x_{t-1} - \mu_{t-1})^2 \right)^{1/2}} \quad (4.19)$$

Teorema 4.1. Bajo el supuesto 1, si $\text{Sup}_t E|u_t|^{2\beta} < \infty$ para $\beta > 2$ y si $l = O(T^{1/4})$ entonces bajo $H_0: a(1) = 0$,

$$Z_{\hat{a}(1)} \text{ converge en distribución a } \frac{\left(\frac{1}{2}\right)(\beta(1)^2 - 1)}{\int_0^1 B(r)^2 dr} \quad (4.20)$$

$$Z_{t_{\hat{a}(1)}} \text{ converge en distribución a } \frac{\left(\frac{1}{2}\right)(B(1)^2 - 1)}{\left[\int_0^1 B(r)^2 dr \right]^{1/2}} \quad (4.21)$$

Nótese que ahora las distribuciones límite de $Z_{\hat{a}(1)}$ y de $Z_{t_{\hat{a}(1)}}$ no dependen de parámetros que no sean de interés. De hecho estas distribuciones límite corresponden a las distribuciones límite de los contrastes de Dickey-Fuller (1979). A

su vez ésta es la distribución límite en el caso en que $\sigma_u^2 = \sigma^2$. Esta circunstancia se da si u_t es i.i.d $(0, \sigma^2)$.

A continuación vamos a resaltar la diferencia entre el contraste de Dickey y Fuller (1979) y el de Phillips (1987). Dickey y Fuller proponen introducir tantos retardos de la variable dependiente como sean necesarias para conseguir que el término de error, u_t , sea i.i.d $(0, \sigma^2)$. A este contraste se le denomina el contraste aumentado de Dickey y Fuller (*ADF*). Phillips, por el contrario, propone una corrección no paramétrica de $\hat{a}(1)$ y de su *t*-ratio de forma que su distribución asintótica coincida con la de u_t cuando es i.i.d $(0, \sigma^2)$, ecuaciones (4.18) y (4.19).

El mayor problema de estos contrastes de raíces unitarias es su *falta de poder* frente a alternativas con raíces menores que la unidad pero próximas a uno, cuando el tamaño muestral es pequeño. Los resultados dependen del *tamaño de la raíz unitaria*, véase Cochrane (1986), así como de la *amplitud del período histórico* que recorren los datos. El obtener muchas observaciones de un período corto de tiempo no resuelve el problema, véase Shiller y Perron (1985). Es incluso mejor tener menos observaciones y cubrir un mayor período de tiempo. Reescribiendo la ecuación (4.3) sin desviaciones con respecto a la media.

$$(1 - B)x_t = (1 - B)\mu_t + a(1)\mu_{t-1} - a(1)x_{t-1} + u_t \quad (4.22)$$

El procedimiento para contrastar $H_0: a(1) = 0$ puede ser el siguiente. Estimar el modelo bajo H_0 ,

$$(1 - B)x_t = (1 - B)\mu_t + u_t \quad (4.23)$$

Si $(1 - B)\mu_t$ es una constante y/o una tendencia y es significativa en la regresión (4.23) entonces, al llevar a cabo el contraste en (4.22), el *t*-ratio del coeficiente de la variable x_{t-1} sigue asintóticamente una distribución normal, ver West (1988). En caso contrario el *t*-ratio seguirá la distribución (4.12). Si u_t es i.i.d $(0, \sigma_u^2)$ entonces el *t*-ratio de x_{t-1} seguirá la distribución (4.21) cuyos valores críticos aparecen tabulados en Fuller (1976). Estos valores críticos varían según sea $(1 - B)\mu_t = 0$, $(1 - B)\mu_t = \text{constante}$, o incluso si hay una tendencia incluida en μ_{t-1} , ecuación (4.22). Se recomienda proceder *de lo general a lo particular*, tanto a la hora de identificar el número, d , de raíces unitarias, como a la hora de determinar si deben aparecer μ_{t-1} y/o $(1 - B)\mu_t$ en la regresión (4.22).

5. Discusión sobre los conceptos de integración y cointegración

La metodología de análisis univariante de series temporales descrita en Box-Jenkins (1976) dio gran popularidad a los modelos denominados ARIMA (p, d, q) , autorregresivos, integrados, de medias móviles. El orden de integración, d , indica el número de veces que hay que diferenciar la serie antes de pasar a identificar el orden, p , de la parte autorregresiva y el orden, q , de la media móvil.

Engle y Granger (1990) resumen estas ideas en la siguiente definición.

Definición 5.1

Una serie temporal sin componentes determinísticos que tiene una representación ARMA estacionaria e invertible después de diferenciarla d veces se dice que es *integrable de orden d* , $x_t \sim I(d)$.

Cuando x_t tiene componentes determinísticos siempre se pueden sustraer incorporándolos en la media μ_t . De forma que x_t es $I(d)$ si $(1 - B)^d(x_t - \mu_t)$ tiene una representación ARMA estacionaria e invertible.

Esta es una definición precisa y útil dentro del contexto de los modelos lineales, donde la única fuente de no estacionariedad de $(x_t - \mu_t)$ es la existencia de raíces unitarias. Si se quieren extender estos conceptos a modelos con incertidumbre, donde la varianza incondicional no es constante, la anterior definición es demasiado restrictiva.

La idea se entiende fácilmente con un ejemplo. Sea x_t el precio de una acción en un mercado financiero eficiente sujeto a variaciones en la incertidumbre del mercado. Si el mercado es eficiente, el cambio en el precio de la acción debería ser impredecible para los agentes que actúan en el mercado. El precio x_t vendrá generado por

$$\Delta x_t = u_t \quad (5.1)$$

donde u_t es independiente con media cero y varianza cambiante (heteroscedasticidad) que capta la incertidumbre del mercado.

Si utilizamos la definición 5.1 de integrabilidad, el modelo (5.1) no puede ser estudiada dentro de la clase de modelos ARIMA(p, d, q) ya que a pesar de tomar una diferencia Δx_t no es estacionario.

Dentro del contexto de los modelos no lineales interesa estudiar las propiedades de transformaciones no lineales de variables integradas. Para ello necesitamos saber las propiedades de transformaciones no lineales de modelos ARIMA, y esto es difícil de tratar.

Para resolver estas lagunas, en Escribano (1987, a) se propone un concepto de integración que no depende de una representación paramétrica de tipo lineal, ni de que la única fuente de no estacionariedad sea la existencia de raíces unitarias.

La idea se basa en hacer explícita la implicación fundamental de que las variables sean $I(1)$. Como vimos en la sección 2, las raíces unitarias generan tendencias en la varianza del proceso. De esta forma podemos encontrar que, si la serie x_t tiene heteroscedasticidad pero no una tendencia en la varianza, entonces es posible que sea $I(0)$ aunque no sea estacionaria.

Antes de pasar a la definición formal de integrabilidad vamos a introducir algunos conceptos que nos serán útiles.

Diremos que x_t es *estacionario en sentido débil* (o de segundo orden) si su media es constante y la covarianza entre x_t y $x_{t+\tau}$ depende sólo de τ y no de t , $\text{cov}(x_t, x_{t+\tau}) = \gamma(\tau)$.

En la metodología de Box-Jenkins (1976), la identificación del número de raíces unitarias se hace en base al *autocorrelograma* de la serie x_t y de sus diferencias. Si x_t es $I(1)$ sus autocorrelaciones teóricas serán igual a uno para

cualquier desfase τ . En la práctica se identifica por un decaimiento lento (menor que exponencial) del correlograma, empezando en un valor cercano a la unidad.

Denominemos a la correlación entre x_t y $x_{t+\tau}$ por $\rho_x(t, \tau)$. Si $\rho_x(t, \tau) \rightarrow 0$, cuando $\tau \rightarrow \infty$, a una tasa exponencial, entonces será $I(0)$ y a su vez será sumable en términos absolutos $\sum_{\tau=0}^{\infty} |\rho_x(t, \tau)| < \infty$. Esta condición garantiza que el espectro normalizado es finito en la frecuencia cero (la frecuencia que indica el largo plazo).

Definición 5.2

x_t es *asintóticamente incorrelacionado en sentido fuerte* si:

- i) $0 < \text{var}(x_t) < \infty \quad \forall t$
- ii) existe una función ρ_τ para $\tau \neq 0$ tal que $\sum_{\tau=0}^{\infty} |\rho_\tau| < \infty$ y donde $|\rho_x(t, \tau)| \leq |\rho_\tau|$ para todo $\tau \neq 0$.

Si x_t es *asintóticamente incorrelacionado en sentido fuerte*, entonces el autocorrelograma decaerá exponencialmente. Nótese que x_t no tiene que ser estacionario en sentido débil ya que las correlaciones $\rho_x(t, \tau)$ son no sólo función de τ sino también de t .

En esta sección nos vamos a reducir al estudio de tendencia temporales no acotadas. Los interesados en las tendencias acotadas pueden consultar Escribano (1987, a).

Definición 5.3

x_t tiene una *tendencia de orden $h(i, t)$ en el i -ésimo momento* si:

- i) $E(x_t^i)$ es finito para valores finitos de t .
- ii) $E(x_t^i) = h(i, t)$ donde $h(i, t)$ depende del tiempo y $|h(i, t)| \rightarrow \infty$ cuando $t \rightarrow \infty$.

En base a este concepto de tendencia si $x_t = ct + \varepsilon_t$, donde ε_t tiene media cero y varianza constante, x_t tiene tendencia en la media, $E(x_t) = ct$, pero no en la varianza, $\text{var}(x_t) = E(x_t - ct)^2 = \sigma_\varepsilon^2$. Sin embargo, un camino aleatorio, $\Delta x_t = \varepsilon_t$, tiene tendencia únicamente en la varianza $\text{var}(x_t) = t\sigma_\varepsilon^2$, ver sección 2. Con este concepto de tendencia y el de *asintóticamente incorrelacionado* podemos pasar a definir lo que entendemos por variables integradas.

Definición 5.4

x_t es *integrada de orden d en sentido débil, $I(d)$* , si:

- i) $x_t - \mu_t$ tiene una tendencia en la varianza y

- ii) después de diferenciar al menos d -veces $(1 - B)^d(x_t - \mu_t)$:
 - a) no tiene tendencia en varianza
 - b) es asintóticamente incorrelacionada en sentido fuerte, y
 - c) es estacionaria en sentido débil.

Nótese que bajo estas condiciones $(x_t - \mu_t)$ tiene una representación ARIMA (p, d, q) . Si estuviéramos en el caso de que el término de ruido blanco, ε_t , tuviera heterocedasticidad, pero no tendencia en la varianza entonces no se cumpliría la propiedad ii), c) pero sí la ii) a) y b). En ese caso diríamos que x_t es *heterogéneamente integrada de orden d , $HI(d)$, en sentido débil*. Un ejemplo de $HI(1)$ es el modelo (5.1) antes discutido.

Escribano (1987, b) sustituye la condición c) por la de estacionaridad estricta y analiza las propiedades de funciones no lineales de variables $I(d)$ y $HI(d)$, y en base a estos conceptos deriva los modelos de corrección de error no lineales.

La denominación de integración débil ha sustituido a la de integración en varianza de Escribano (1987, a) para evitar que sea confundido con el concepto de integración en varianza introducido por Engle y Bollerslev (1986) en el contexto de los modelos ARCH.

En Escribano (1987, a) se demuestra que las condiciones del supuesto 4.1, de la sección 4, introducidas por Phillips (1987) son suficientes para que la u_t de la ecuación (4.1) sea integrada de orden cero, $I(0)$ en sentido débil. Si relajáramos la condición de que u_t tiene varianza constante, entonces u_t sería heterogéneamente integrada de orden cero, $HI(0)$ en sentido débil.

6. Cointegración e identificación del número de raíces unitarias en modelos multivariantes

Una vez definido el concepto de integración de forma general podemos explicitar lo que se entiende por cointegración y luego pasar a analizar sus implicaciones en términos de los modelos de series temporales multivariantes.

Granger (1981) introdujo el concepto de cointegración para establecer si variables que siguen sendas de crecimiento evolucionan o no de forma paralela. Para una introducción al estudio de las implicaciones de este concepto véase Granger (1986). Sea X_t un vector columna de orden $N \times 1$. Los componentes de este vector son variables económicas y las representaremos de forma genérica por x_{jt} donde $j = 1, 2, \dots, N$. Sea μ_t la media de X_t .

Definición 6.1

Los componentes x_{jt} del vector X_t están *cointegrados en sentido débil* si:

- a) todos los componentes son integrados del mismo orden, $I(d)$ en sentido débil para $d \in \mathbb{N}$, y
- b) existe una combinación lineal de ellos $Z_t^* = \alpha'(X_t - \mu_t)$ que es $I(d - b)$,

$b > 0$, en sentido débil. A la matriz α' de orden $r \times N$ se le denomina matriz de cointegración, $r < N$.

La existencia de raíces unitarias genera tendencias en el momento de segundo orden (la varianza) y quizás también en la media. Sin embargo, puede haber procesos estocásticos que generen tendencia en la media y la varianza pero que no tengan raíces unitarias y, por tanto, no sean procesos integrados. Un concepto más general que el de cointegración es el de cotendencias introducido en Escribano (1987, a).

Definición 6.2

Los componentes x_{jt} del vector X_t tienen *co-tendencias en el i -ésimo momento* si:

- todos los componentes tienen tendencias del mismo orden en el i -ésimo momento y
- existe una combinación lineal de ellos $Z_t = \alpha' X_t$ que o no tiene tendencia en el i -ésimo momento o tiene una tendencia de menor orden que los componentes x_{jt} . α' es una matriz $r \times N$ con $r < N$.

Los casos empíricamente más relevantes son los de co-tendencias en la media y la varianza. Un ejemplo de co-tendencias en media es

$$x_{jt} = c_j + b_j t + \varepsilon_{jt} \quad j = 1, 2 \quad (6.1)$$

Si $\alpha' = (1, -\alpha_1)$ y $X_t' = (x_{1t}, x_{2t})$, la combinación lineal Z_t será

$$Z_t = \alpha' X_t = (1, -\alpha_1) \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = x_{1t} - \alpha_1 x_{2t} \quad (6.2)$$

sustituyendo (6.1) en (6.2) se obtiene

$$Z_t = (c_1 - \alpha_1 c_1) + (b_1 - \alpha_1 b_2)t + (\varepsilon_{1t} - \alpha_1 \varepsilon_{2t}) \quad (6.3)$$

la condición que debe cumplirse para que Z_t no tenga tendencia en la media es que $b_1 - \alpha_1 b_2 = 0$ y en ese caso diremos que los componentes x_{1t} y x_{2t} tienen cotendencias en media.

A continuación vamos a analizar las implicaciones, sobre los modelos multivariantes de seis temporales, del hecho de que sus componentes estén cointegrados y tengan co-tendencias en la media.

Sea $(X_t - \mu_t)$ un vector columna de orden $N \times 1$ de componentes x_{jt} que son $I(d)$ en sentido débil. Nos vamos a centrar en el estudio de los casos empíricamente más relevantes. En primer lugar estudiaremos el caso $d = 1$ y en segundo lugar el $d = 2$.

6.1. Variables $I(1)$ en sentido débil

Si las variables son $I(1)$ en sentido débil, entonces tienen una representación de Wold del tipo

$$(1 - B)(X_t - \mu_t) = \Theta(B)\varepsilon_t \quad (6.1.1)$$

donde $\Theta(B)$ es una matriz de orden $N \times N$, de polinomios en el operador de retardos B . Las raíces de $|\Theta(B)| = 0$ están fuera del círculo de la unidad. ε_t es un vector de términos serialmente incorrelacionados con media cero y matriz de varianzas-covarianzas igual a Σ .

Siguiendo los pasos descritos en la sección 2, pero aplicados al caso multivariante, podemos hacer una *expansión de Taylor* alrededor del punto $B = 1$

$$\Theta(B) = \Theta(1) + (1 - B)\Theta^*(B) \quad (6.1.2)$$

donde la matriz $\Theta^*(B) \equiv \frac{\partial \Theta(B)}{\partial B}$ cumple que $|\Theta^*(B)| = 0$ tiene todas las raíces fuera del círculo de la unidad.

Sustituyendo (6.1.2) en (6.1.1) se obtiene

$$(1 - B)(X_t - \mu_t) = \Theta(1)\varepsilon_t + (1 - B)\Theta^*(B)\varepsilon_t \quad (6.1.3)$$

Imponiendo las condiciones iniciales de que $x_{ji} = 0$ para todo j y para todo $i \leq 0$, podemos escribir (6.1.3) en forma extendida dividiendo por $(1 - B)$,

$$(X_t - \mu_t) = \Theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + \Theta^*(B)\varepsilon_t \quad (6.1.4)$$

Escribiendo (6.1.4) sin desviaciones con respecto a la media se obtiene

$$X_t = \mu_t + \Theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + \Theta^*(B)\varepsilon_t \quad (6.1.5)$$

Si denominamos $\tau_t = \mu_t + \Theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j}$ y $\mu_t = bt$ entonces podemos escribir (6.1.5) como

$$X_t = \tau_t + \Theta^*(B)\varepsilon_t \quad (6.1.6.a)$$

$$\tau_t = b + \tau_{t-1} + \eta_t \quad (6.1.6.b)$$

$$\eta_t = \Theta(1)\varepsilon_t \quad (6.1.6.c)$$

Nótese que esta representación es una generalización de la representación de Beveridge y Nelson (1981), obtenida en la sección 2, al caso multivariante, véase

Escribano (1987, a). La ecuación (6.1.6.b) es una representación de tendencias estocásticas para el caso en que τ_t es un vector de orden $N \times 1$.

Si los componentes de $(X_t - \mu_t)$ están cointegrados en sentido débil de forma que $\alpha'(X_t - \mu_t)$ sea $I(0)$ en sentido débil, vamos a analizar cuáles son las restricciones que se imponen sobre los parámetros de la ecuación (6.1.5). Premultiplicando en (6.1.5) por la matriz α' de orden $r \times N$, se obtiene que

$$\alpha'X_t = \alpha'\mu_t + \alpha'\Theta(1) \sum_{j=0}^{t-1} \varepsilon_{t-j} + \alpha'\Theta^*(B)\varepsilon_t \quad (6.1.7)$$

Para que $\alpha'X_t$ sea $I(0)$ en sentido débil debe cumplirse que $\alpha'\Theta(1) = 0$. Si éste es el caso, entonces los elementos de las columnas de $\Theta(1)$ serán linealmente dependientes y, por tanto, $\Theta(1)$ será de rango $N - r$. Si los componentes de la media de X_t tienen tendencias temporales, entonces para que los componentes de X_t tengan *co-tendencias en la media* se debe cumplir que $\alpha'\mu_t = \text{constante}$. Bajo estas condiciones

$$\alpha'X_t = \alpha'\Theta^*(B)\varepsilon_t \quad (6.1.8)$$

Si $\alpha'\mu_t = 0$ y $\alpha'\Theta(1) = 0$, entonces en la representación multivariante de las ecuaciones (6.1.6.a)-(6.1.6.c) se debe cumplir que $\alpha'\tau_t = 0$ y que $\alpha'\eta_t = 0$.

El hecho de que $\alpha'\tau_t = 0$ implica que existen *co-tendencias* entre las tendencias estocásticas de los componentes x_{jt} . Esta idea llevó a Stock y Watson (1988) a proponer una *representación de tendencias comunes*, o de factores comunes en la terminología de Peña (1990).

Para obtener la representación de tendencias comunes de Stock y Watson se puede descomponer la matriz $\Theta(1)$ en la forma canónica de Jordán.

$$\Theta(1) = HJH^{-1} = (H_1, H_2) \begin{pmatrix} J_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \begin{pmatrix} H^1 \\ H^2 \end{pmatrix} = H_1 J_{11} H^1 \quad (6.1.9)$$

donde los órdenes de las matrices son los siguientes: $\Theta(1)$ es $N \times N$, H_1 es $N \times (N - r)$, H_2 es $N \times r$, J_{11} es $(N - r) \times (N - r)$, O_{22} es $r \times r$, H^1 es $(N - r) \times N$ y H^2 es $r \times N$.

Las columnas de H son los vectores propios de $\Theta(1)$. La matriz J diagonal por bloques, tiene los valores propios de $\Theta(1)$ en la diagonal principal.

Si $\alpha'\Theta(1) = 0$, entonces $\alpha'H_1 J_{11} H^1 = 0$ debido a que $\alpha'H_1 = 0$. Por tanto, los vectores propios (autovectores) son linealmente dependientes.

Utilizando la descomposición (6.1.9) en (6.1.5), obtenemos

$$X_t = \mu_t + H_1 J_{11} H^1 \sum_{j=0}^{t-1} \varepsilon_{t-j} + \Theta^*(B)\varepsilon_t \quad (6.1.10)$$

Si $\mu_t = bt$ y $\alpha'\mu_t = \alpha'bt = 0$, entonces $b = H_1 b^*$. Sustituyendo en (6.1.10)

$$X_t = H_1 \left(b^* t + J_{11} H^1 \sum_{j=0}^{t-1} \varepsilon_{t-j} \right) + \Theta^*(B)\varepsilon_t \quad (6.1.11)$$

Denominando $\tau_t^* = b^*t + J_{11}H^1 \sum_{j=0}^{t-1} E_{t-j}$ obtenemos la siguiente *representación de tendencias comunes*.

$$X_t = H_1 \tau_t^* + \Theta^*(B)\varepsilon_t \quad (6.1.12.a)$$

$$\tau_t^* = b + \tau_{t-1}^* + \eta_t^* \quad (6.1.12.b)$$

$$\eta_t^* = J_{11}H^1 \varepsilon_t \quad (6.1.12.c)$$

donde τ_t^* es de orden $(N-r) \times 1$ y η_t^* es $(N-r) \times 1$.

Comparando las ecuaciones de las expresiones (6.1.6) y (6.1.12), observamos que mientras en (6.1.6) si no hay cointegración, τ_t es un vector de orden $N \times 1$, si se da la cointegración la nueva tendencia τ_t^* se reduce a un vector de $(N-r) \times 1$ ya que hay r co-tendencias en media y r relaciones de cointegración, o, lo que es lo mismo, habrá $(N-r)$ *tendencias comunes* o *factores comunes*, ver Stock y Watson (1988), King *et. al.* (1987), Peña (1990) y Fernández *et. al.* (1987).

La representación de tendencias comunes (6.1.12) impone que la matriz α' de cointegración débil sea a la vez la matriz que anule las tendencias de la media (co-tendencias). Si éste no fuera el caso, de (6.1.7) obtendríamos que

$$\alpha'X_t = \alpha'\mu_t + \alpha'\Theta^*(B)\varepsilon_t \quad (6.1.13)$$

de forma que la relación de cointegración deberá incluir, como variable explicativa, la tendencia determinística $\alpha'\mu_t$.

En la expresión (6.1.12), debido a la cointegración débil y a las co-tendencias en media, hay menos tendencias, τ_t^* , que en la versión (6.1.6), τ_t . Como estas tendencias estocásticas han sido generadas por las raíces unitarias de (6.1.1) quiere decir que si los elementos del vector $(X_t - \mu_t)$ están cointegrados, entonces no habrá N raíces unitarias sino que habrá $(N-r)$. Por tanto, si diferenciamos las series individuales en el modelo multivariante, estaremos sobrediferenciando y perdiendo la información en común sobre el largo plazo de las series. Como hemos visto, la representación de medias móviles es muy útil para analizar las condiciones que han de cumplirse para que las variables estén co-integradas y tengan co-tendencias en media. Sin embargo, a la hora de estimar modelos multivariantes, la representación de vectores autorregresivos (VAR) es más útil. Premultiplicando en la ecuación (6.1.1) por $\Theta(B)^{-1}$ obtenemos la representación vectorial autorregresiva posiblemente de orden infinito,

$$A(B)(X_t - \mu_t) = \varepsilon_t \quad (6.1.14)$$

donde $A(B)$ es una matriz de orden $N \times N$ cuyos elementos son polinomios en el operador de retardo B . Haciendo una expansión de Taylor alrededor del punto $B = 1$

$$A(B) = A(1) + \frac{\partial A(B)}{\partial B}(1-B) = A(1) + A^+(B)(1-B) \quad (6.1.15)$$

Sustituyendo (6.1.15) en (6.1.14) y reordenando términos obtenemos la *representación de Bewley*, ver Bewley (1979)

$$A(1)(X_t - \mu_t) = -A^+(B)(1 - B)(X_t - \mu_t) + \varepsilon_t \quad (6.1.16)$$

Sumando y restando a (6.1.15) la matriz $A(1)B$

$$A(B) = A(1)B + [A(1) + A^+(B)](1 - B) = A(1)B + A^*(B)(1 - B) \quad (6.1.17)$$

Sustituyendo (6.1.17) en (6.1.14) obtenemos la *representación de corrección de error*, ver Engle y Granger (1990)

$$A^*(B)(1 - B)(X_t - \mu_t) = -A(1)(X_{t-1} - \mu_{t-1}) + \varepsilon_t \quad (6.1.18)$$

Nótese que tanto $A^+(B)$ en la representación de Bewley como $A^*(B)$ en la representación de corrección de error tendrán las raíces de sus ecuaciones características fuera del círculo de la unidad, dado que los $(X_t - \mu_t)$ son $I(1)$ y ya se han aislado las correspondientes raíces unitarias en el término $(1 - B)$.

Es importante resaltar que la matriz que recoge los efectos de largo plazo $A(1)$; aparece de forma explícita tanto en la formulación de Bewley como en la de corrección de error. Las ventajas y desventajas de utilizar una u otra representación serán discutidos en la siguiente sección.

Nótese la similitud que hay entre los modelos (3.7) y los (6.1.18). De forma inmediata se observa que realizar el contraste sobre si los términos de corrección de error, $A(1)(X_{t-1} - \mu_{t-1})$, son significativos equivale a nivel univariante a realizar el contraste de Dickey-Fuller. Se concluye por consiguiente, que el obtener un término de corrección de error significativo equivale a decir que en el modelo de N variables al modelizarlas de forma conjunta existe un número de raíces unitarias menor que N . Por tanto, la representación VAR con variables diferenciadas estará mal especificada, en el sentido de que estaremos sobrediferenciando las series.

La representación de corrección de error, ecuación (6.1.18), se puede escribir sin desviaciones con respecto a la media de la siguiente forma.

$$A^*(B)(1 - B)X_t = A^*(B)(1 - B)\mu_t + A(1)\mu_{t-1} - A(1)X_{t-1} + \varepsilon_t \quad (6.1.19)$$

los términos de X_t son $I(1)$ y, por tanto $(1 - B)X_t$ será $I(0)$. Para que este modelo no esté mal especificado es necesario que $A(1)X_{t-1}$ sea $I(0)$ y, por tanto, los componentes de X_t han de estar *cointegrados en sentido débil*. A su vez concluimos que si μ_t tiene tendencias determinísticas y $A^*(B)(1 - B)\mu_t + A(1)\mu_{t-1} = C$ de orden $N \times 1$, la expresión (6.1.19) se convierte en

$$A^*(B)(1 - B)X_t = C - A(1)X_{t-1} + \varepsilon_t \quad (6.1.20)$$

La *tendencia* de μ_t debe ser *lineal*, de forma que $(1 - B)\mu_t$ no tenga tendencia, y los términos de μ_t deben tener *co-tendencias*, de forma que $A(1)\mu_{t-1}$ no tenga tendencia.

De la discusión basada en las ecuaciones (6.1.19) y (6.1.20) se deriva el *teorema de representación de Granger* que prueba que los componentes del vector X_t están cointegrados si y sólo si existe un modelo de corrección de error, ver Engle y Granger (1990).

6.2. Descomposición de Smith-McMillan-Yoo

Yoo (1986) extendió el estudio de los modelos de corrección de error basados en variables que son $I(1)$ al caso general de variables $I(d)$. Para ello se basó en la descomposición de Smith-McMillan, ver Kailath (1980), y generalizó el concepto de cointegración al de cointegración polinomial. En este apartado se pretende hacer una introducción a este procedimiento alternativo. Una aplicación al caso de frecuencias estacionales se encuentra en Hylleberg *et. al.* (1990). Partimos de la representación ARMA (p, q) multivariante, aplicada a variables que son $I(1)$ en sentido débil,

$$(1 - B)X_t^* = \Theta(B)\varepsilon_t \quad (6.2.1)$$

donde la matriz $\Theta(B)$ es una matriz racional de orden $N \times N$ cuyos elementos son polinómicos en el operador de retardos B . X_t^* indica que los componentes de X_t están en desviaciones con respecto a la media. $X_t^* = X_t - \mu_t$. El siguiente lema, cuya prueba se encuentra en Yoo (1986), nos va a servir para desarrollar las diversas representaciones de forma rigurosa.

Lema 1. De cualquier matriz racional $\Theta(B)$, que sea finita para todo B dentro del círculo de la unidad o en el círculo de la unidad, se puede obtener una matriz $M(B)$,

$$M(B) = U(B)\Theta(B)V(B) \quad (6.2.2)$$

donde

- 1) $|U(B)| = 0$ y $|V(B)| = 0$ tienen todas las raíces fuera del círculo de la unidad,
- 2) $M(B)$ es una matriz diagonal y las raíces de $|M(B)| = 0$ están en el círculo de la unidad o dentro del círculo de la unidad.

Utilizando la ecuación (6.2.2) de este lema podemos escribir la matriz (6.2.1) de $\Theta(B)$ como

$$\Theta(B) = U(B)^{-1}M(B)V(B)^{-1} \quad (6.2.3)$$

ya que las inversas de $U(B)$ y de $V(B)$ existen por la condición 1) del lema 1. Por la condición 2) del lema 1 sabemos que las raíces unitarias de X_t^* dependerán de las raíces unitarias que haya en $M(B)$. Engle y Yoo (1989) hacen explícita esta idea en la siguiente definición

Definición 6.3

El sistema (6.2.1), es *integrado de orden uno*, $I(1)$, si la máxima potencia de $(1 - B)$ es $M(B)$ es uno y la mínima potencia es cero.

La idea detrás de esta definición queda clara si sustituimos la expresión (6.2.3) en (6.2.1),

$$(1 - B)X_t^* = U(B)^{-1}M(B)V(B)^{-1}\varepsilon_t \quad (6.2.4)$$

Si la matriz $M(B)$ tiene algún elemento en su diagonal principal igual a $(1 - B)$ entonces se cancelará con el término $(1 - B)$ de la izquierda de la ecuación y por tanto ese componente será $I(0)$ en vez de $I(1)$. A continuación vamos a analizar esta cuestión más en detalle.

De la ecuación (6.2.3) se observa que $\Theta(1) = U(1)^{-1}M(1)V(1)^{-1}$. El rango de $\Theta(1)$ será el mismo que el de $M(1)$. Si ninguna de las variables de X_t^* están cointegradas el rango será N . Si hay r relaciones de cointegración entonces el rango de $\Theta(1)$ y de $M(1)$ será $N - r$. En este caso la matriz $M(B)$ tendrá la forma siguiente,

$$M(B) = \begin{bmatrix} I_{N-r} & 0 \\ 0 & (1 - B)I_r \end{bmatrix} \quad (6.2.5)$$

donde I_r indica la matriz identidad de orden $r \times r$.

Para obtener la representación autorregresiva, VAR, seguiremos los siguientes pasos. Primero, multiplicamos por $U(B)$ en (6.2.4)

$$U(B)(1 - B)X_t^* = M(B)V(B)^{-1}\varepsilon_t \quad (6.2.6)$$

Definamos $\bar{M}(B)$ de forma que $\bar{M}(B)M(B) = (1 - B)I_{N \times N}$,

$$\bar{M}(B) = \begin{bmatrix} (1 - B)I_{N-r} & 0 \\ 0 & I_r \end{bmatrix} \quad (6.2.7)$$

Segundo, premultiplicando por $\bar{M}(B)$ en (6.2.6)

$$\bar{M}(B)U(B)X_t^* = V(B)^{-1}\varepsilon_t \quad (6.2.8)$$

Tercero, premultiplicando por $V(B)$ se obtiene la representación VAR

$$A(B)X_t^* = \varepsilon_t \quad (6.2.9)$$

donde $A(B) = V(B)\bar{M}(B)U(B)$. Nótese que $A(B)$ sólo tiene $N - r$ raíces unitarias ya que sólo hay $N - r$ raíces unitarias en $\bar{M}(B)$.

Para obtener una representación de corrección de error se puede realizar una expansión de Taylor de $\bar{M}(B)$ alrededor del punto $B = 1$.

$$\bar{M}(B) = \bar{M}(1) + (1 - B)\bar{M}(B)^* = \begin{bmatrix} O_{N-r} & 0 \\ 0 & I_r \end{bmatrix} + (1 - B)\begin{bmatrix} I_{N-r} & 0 \\ 0 & O_r \end{bmatrix} \quad (6.2.10)$$

Sumando y restando $\bar{M}(1)B$ en (6.2.10) se obtiene

$$\bar{M}(B) = \begin{bmatrix} O_{N-r} & 0 \\ 0 & I_r \end{bmatrix} B + (1 - B) \begin{bmatrix} I_{N-r} & 0 \\ 0 & I_r \end{bmatrix} \quad (6.3.11)$$

Sustituyendo (6.3.11) en (6.2.9) y particionando $V(B) = [V_1(B), \gamma(B)]$ y $u'(B) = [u_1(B), \alpha(B)]$ y sustituyendo en la matriz $A(B)$ se obtiene la siguiente reparametrización de $A(B)$,

$$A(B) = [V_1(B)U_1(B) + \gamma(B)\alpha'(B)](1 - B) + \gamma(B)\alpha'(B)B \quad (6.3.12)$$

Sustituyendo (6.3.12) en (6.2.9) se obtiene la *representación de corrección de error* siguiente,

$$[V_1(B)U_1(B) + \gamma(B)\alpha'(B)](1 - B)X_t^* = -\gamma(B)\alpha'(B)X_{t-1}^* + \varepsilon_t \quad (6.3.13)$$

o de forma equivalente

$$V(B)U(B)(1 - B)X_t^* = -\gamma(B)\alpha'(B)X_{t-1}^* + \varepsilon_t \quad (6.3.14)$$

En este caso $\alpha'(B)X_t^*$ es $I(0)$ y como los elementos de X_t^* son $I(1)$ estamos frente a un caso de *cointegración polinómica*, ya que la matriz de cointegración $\alpha'(B)$ es una matriz de polinomios en el operador de retardos B . Esta generalización del concepto de cointegración fue introducida por Yoo (1986). Los ajustes hacia el equilibrio a largo plazo tampoco son constantes, $\gamma(B)$, y por tanto habrá retardos distribuidos polinómicos en los términos de corrección de error.

Cabe preguntarse cuál es la relación entre esta representación de corrección de error y la obtenida en la sección anterior, ecuación (6.1.18).

La matriz $M(B)$ está relacionada con la matriz $\Theta(B)$ de forma única. Sin embargo las matrices $V(B)$ y $U(B)$ no son únicas. Por tanto, cabría pensar que es posible encontrar parametrizaciones de $V(B) = [V_1^*(B), \gamma]$ y de $U'(B) = [U_1^*(B), \alpha]$ cuyas matrices γ y α sean constantes. Yoo (1986) demuestra que no siempre es posible encontrar matrices de cointegración que sean constantes, ver también Engle y Yoo (1989).

Casos de cointegración polinomial se dan cuando las variables son, por ejemplo, $I(2)$. En Andrés *et. al.* (1990), en Dolado (1989) y en Engle y Yoo (1989), se derivan modelos de corrección de error para este caso. En Granger y Lee (1989) se considera un caso particular de cointegración polinómica denominado *multicointegración*.

7. Inferencia y contrastes en modelos con variables cointegradas

En la sección anterior hemos visto las distintas representaciones de modelos multivariantes cuyas variables están cointegradas. Si X_t^* es un vector de N componentes, la matriz de cointegración α' es una matriz de orden $r \times N$ y por tanto el rango de cointegración es $r < N$. Que el rango de cointegración sea r

quiere decir que hay r vectores de cointegración entre los N variables que forman el vector X_t^* .

En la representación de medias móviles vectoriales obtenida en (6.1.5), vimos que cuando hay r relaciones de cointegración el rango de la matriz $\Theta(1)$ debería ser $N - r$. Stock y Watson (1988) demostraron que en ese caso existe una representación con $N - r$ tendencias comunes, ver ecuación (6.1.12).

En esta sección vamos a estudiar cómo *identificar el rango de cointegración* r . Para ello vamos a considerar primero el caso en donde $r = 1$ y luego donde $r \geq 1$. Identificar el rango de cointegración r es equivalente a encontrar el número $N - r$ de raíces unitarias, o tendencias comunes, que hay en el modelo multivariante.

7.1. El rango de cointegración es $r = 1$

Por sencillez en la exposición vamos a considerar que el orden de integración débil es uno, $I(1)$, y por tanto cada componente de X_t^* es $I(1)$. En este caso $\alpha' = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$, es un vector de orden $N \times 1$. Para estimar el vector α' primero normalizamos de forma arbitraria, por ejemplo dividiendo por α_1^* ,

$$x_{t1} = \alpha_2 x_{t2} + \alpha_3 x_{t3} + \dots + \alpha_N x_{tN} + Z_t \quad (7.1.1)$$

donde los coeficientes $\alpha_i = \frac{\alpha_i^*}{\alpha_1^*}$. La variable Z_t recoge los errores con respecto a la situación de equilibrio a largo plazo y será $I(0)$ si las variables están cointegradas. Nótese que no todos los N componentes de X_t^* deben formar parte de (7.1.1) ya que es posible que algunos $\alpha_i = 0$. Z_t es en la nomenclatura de la sección anterior igual a $\alpha' X_t^*$. Estos términos de la relación de cointegración entran en la representación de corrección de error a través del término $A(1)X_{t-1}^*$, ver ecuación (6.1.18), donde la matriz $A(1)$ se descompone en el producto $A(1) = \gamma\alpha'$ donde γ es una matriz de orden $N \times r$ y α' es $r \times N$. El hecho de poder normalizar la relación de cointegración como en (7.1.1) llevó a Engle y Granger (1990) a proponer la estimación de modelos de corrección de error en 2-etapas. En la primera se estiman los componentes de $\hat{\alpha}'$ por mínimos cuadrados ordinarios y en la segunda se estima el modelo de corrección de error con la variable explicativa $\gamma\hat{Z}_{t-1} = \gamma\hat{\alpha}'X_{t-1}^*$, en vez de γZ_{t-1} .

La justificación de este procedimiento en 2-etapas está basado en la estimación «*superconsistente*» de los parámetros α_i en (7.1.1) cuando las variables están cointegradas, ver Stock (1987). A continuación vamos a derivar este resultado fundamental solamente para el caso de $N = 2$. La formulación más general se encuentra en Andrews (1988). La expresión (7.1.1) se reduce a,

$$x_{t1} = \alpha_2 x_{t2} + z_t \quad (7.1.2)$$

Por simplicidad supondremos que x_{t2} es el proceso $I(1)$ más sencillo, un camino aleatorio sin deriva,

$$\Delta x_{t2} = \varepsilon_t \quad (7.1.3)$$

Los componentes del vector $\begin{pmatrix} z_t \\ \varepsilon_t \end{pmatrix}$ son independientes e idénticamente distribuidos con media cero y matriz de varianzas covarianzas Σ . Supondremos que Σ es una matriz diagonal de forma que no haya correlación contemporánea entre ε_t y z_t .

Si estimamos el parámetro α_2 por mínimos cuadrados ordinarios en la regresión (7.1.2) obtenemos,

$$\hat{\alpha}_2 = \frac{\sum_{t=1}^T x_{t2} x_{t1}}{\sum_{t=1}^T x_{t2}^2} \quad (7.1.4)$$

Sustituyendo x_{t1} en (7.1.4) por el modelo generador (7.1.2) y operando se obtiene

$$\hat{\alpha}_2 = \alpha_2 + \frac{\sum_{t=1}^T x_{t2} \varepsilon_t}{\sum_{t=1}^T x_{t2}^2} \quad (7.1.5)$$

Como por la ecuación (7.1.3) x_{t2} se puede expresar como la suma parcial de los términos ε_t cuando $x_{02} = 0$,

$$x_{T2} = \sum_{t=1}^T \varepsilon_t \quad (7.1.6)$$

Siguiendo los resultados del lema 4.3 de la sección 4 y a Phillips (1987) se obtiene,

$$T^{-2} \sum_{t=1}^T x_{t2}^2 \text{ converge en distribución a } \sigma_2^2 \int_0^1 B_2(r)^2 dr \quad (7.1.7)$$

y

$$T^{-1} \sum_{t=1}^T x_{t2} \varepsilon_t \text{ converge en distribución a } \sigma_2 \sigma_1 \int_0^1 B_2(r) dB_1(r) \quad (7.1.8)$$

donde $B_i(r)$, $i = 1, 2$ son movimientos Brownianos, ver sección 4.

Restando α_2 en (7.1.5) y multiplicando por T se obtiene

$$T(\hat{\alpha}_2 - \alpha_2) = \frac{T^{-1} \sum_{t=1}^T x_{t2} \varepsilon_t}{T^{-2} \sum_{t=1}^T x_{t2}^2} \quad (7.1.9)$$

Utilizando los resultados de (7.1.7) y (7.1.8) obtenemos que

$$T(\hat{\alpha}_2 - \alpha_2) \text{ converge en distribución a } \left(\frac{\sigma_1}{\sigma_2} \right) \frac{\int_0^1 B_2(r) dB_1(r)}{\int_0^1 B_2(r)^2 dr} \quad (7.1.10)$$

Esto es, $T(\hat{\alpha}_2 - \alpha_2)$ converge a un funcional de un movimiento Browniano. Al multiplicar $(\hat{\alpha}_2 - \alpha_2)$ por T en vez de por la normalización clásica $T^{1/2}$, se obtiene que $\hat{\alpha}_2$ converge a α_2 a la tasa $Op(T^{-1})$ y por ello se le denomina el estimador «superconsistente».

Las simulaciones de Monte Carlo realizados por Banerjee, *et. al. ...* (1986) indican que a pesar de la superconsistencia del estimador $\hat{\alpha}_2$, en pequeñas muestras puede haber grandes sesgos. Unas excelentes discusiones de las ventajas comparativas de estimar el vector de cointegración en el modelo de corrección de error en vez de en el modelo estático, ecuación (7.1.2), se encuentran en Gonzalo (1989) y en Dolado, *et al. ...* (1990). Sin embargo, todos estos análisis consideran que la media de X_t es cero. En Escribano (1990) y en Andrés *et. al. ...* (1990) se comentan las ventajas de comparar los resultados de los dos procedimientos cuando existen tendencias en la media.

Para contrastar si la ecuación (7.1.1) es una relación de cointegración hay que analizar si Z_t es $I(0)$. Engle y Granger (1990) proponen contrastar la hipótesis nula de no-cointegración, $Z_t \sim I(1)$, frente a la alternativa $Z_t \sim I(0)$. La hipótesis nula es que Z_t tiene una raíz unitaria. Los contrastes de hipótesis en este caso son similares a los discutidos en las secciones 3 y 4, los valores críticos de los test de Dickey-Fuller son distintos a los obtenidos por Dickey-Fuller en Fuller (1976). Los valores críticos fueron generados por métodos de simulación de Monte Carlo y se encuentran en Engle y Granger (1990) para $N = 2$ y en Engle y Yoo (1987) para $N \geq 2$. Véase Escribano (1986, b) y Andrés *et. al. ...* (1990) para una aplicación.

7.2. El rango de cointegración es $r \geq 1$

Cuando el rango de cointegración es mayor que 1 entonces habrá varios vectores de cointegración. En este caso el procedimiento de estimación en 2-etapas discutido en la sección anterior no parece ser el más apropiado. Sin embargo Engle y Yoo (1989) proponen una generalización de la normalización adoptada en (7.1.1) para el caso matricial. Si la matriz de cointegración es α' de orden $r \times N$, proponen particionar α' en dos matrices α'_1 y α'_2 , donde α'_1 es de orden $r \times r$ y α'_2 de orden $(N - r) \times r$. Eligiendo la matriz $p = \alpha_1^{-1}$, tenemos que $p\alpha'x_t^* = (I_r, \alpha_1^{-1}\alpha'_2)x_t^*$. Los estimadores de mínimos cuadrados ordinarios de cada ecuación son superconsistentes y por tanto justifican el utilizar el método de 2-etapas para el caso de r vectores cointegradores.

Stock y Watson (1988) proponen el siguiente procedimiento para contrastar el rango de cointegración r . Si hay r vectores cointegradores habrá sólo $N - r$ raíces unitarias en la representación vectorial autorregresiva. Para contrastar

que hay $N - r$ raíces unitarias y no $N - (r + 1)$, Stock y Watson proponen seleccionar los $N - r$ componentes principales de X_t^* , asociados a los mayores $N - r$ valores propios, denominados PC_t . El vector PC_t es de orden $(N - r) \times 1$. Si los elementos de PC_t son $I(1)$ entonces la matriz de correlación serial de primer orden de los PC_t tendrá $(N - r)$ raíces unitarias bajo H_0 , y $N - (r + 1)$ bajo la alternativa H_1 . Nótese, que bajo H_0 deberá haber $(N - r)$ valores propios iguales a 1. El contraste por tanto se basará en $T(\hat{\lambda}_{N-(r+1)} - 1)$ donde $\hat{\lambda}_{N-(r+1)}$ es el menor valor propio. Los valores críticos se encuentran en Stock y Watson (1988). Véase la similitud de este procedimiento con el analizado en Peña (1990).

El enfoque que más atención está concentrando cuando $r > 1$ es el propuesto por Johansen (1988). El procedimiento es el de máxima verosimilitud con información completa aplicado al caso de representaciones de corrección de error. En este caso necesitamos incluir el supuesto adicional de que los errores ε_t de la ecuación (6.1.1) siguen una distribución normal multivariante.

El procedimiento sugerido por Johansen (1988) y Johansen y Juselius (1988) es el siguiente.

Se escribe el modelo de corrección de error con el retardo del error en p . Esto se consigue sumando y restando $A(1)B^p$ en vez de $A(1)B$ como en (6.1.17).

Supongamos que $A^*(B)$ es como máximo de orden p . De la ecuación (6.1.18) obtenemos

$$\Delta X_t^* = A_1^* \Delta X_{t-1}^* + \dots + A_p^* \Delta X_{t-p}^* - A(1)X_{t-p}^* + \varepsilon_t \quad (7.2.1)$$

La hipótesis nula, $H_0: A(1) = \gamma\alpha'$ donde γ es de orden $N \times r$, α' es $r \times N$ y α' es la matriz de cointegración de rango r .

Se concentra la función de verosimilitud con respecto a los parámetros $A_1^*, A_2^*, \dots, A_p^*$ mediante las regresiones de ΔX_t^* y de X_{t-p}^* sobre los regresores $\Delta X_{t-1}^*, \Delta X_{t-2}^*, \dots, \Delta X_{t-(p-1)}^*$ y de esta forma generamos los residuos correspondientes R_{ot} y R_{pt} .

Para un α' dado podemos estimar γ en la regresión

$$R_{ot} = \gamma\alpha'R_{pt} + \text{error} \quad (7.2.2)$$

el estimador de γ , como función de α' será

$$\hat{\gamma}(\alpha') = S_{op}\alpha'(S_{pp}\alpha')^{-1} \quad (7.2.3)$$

donde S_{op} y S_{pp} son las matrices de residuos formados como

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it}R'_{jt} \quad i, j = 0, p \quad (7.2.4)$$

calculando los valores propios, λ , en la ecuación

$$|\lambda S_{pp} - S_{po}S_{oo}^{-1}S_{op}| = 0 \quad (7.2.5)$$

determinamos α . Las soluciones $\hat{\lambda}_1 > \dots > \hat{\lambda}_T$ cuyos vectores propios son $\hat{V} = (\hat{V}_1, \dots, \hat{V}_T)$ normalizados por $\hat{V}' S_{pp} \hat{V} = I$. El estimador de máxima verosimilitud de la matriz de cointegración es

$$\hat{\alpha} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r) \quad (7.2.6)$$

Utilizando la normalización anterior en (7.2.3) y sustituyendo (7.2.6) obtenemos que

$$\hat{\gamma} = S_{op} \hat{\alpha} \quad (7.2.7)$$

El contraste de la razón de verosimilitud de $H_0 : \text{rango } A(1) = r$ es,

$$LR \equiv 2 \ln Q = -T \sum_{i=r+1}^T \ln(1 - \hat{\lambda}_i) \quad (7.2.8)$$

los valores críticos de este estadístico obtenidos mediante simulaciones se encuentran en Johansen (1988). Otro estadístico alternativo es el denominado $\lambda_{\text{máx}}$ donde se contrasta $H_0 : \text{rango } A(1) = r - 1$, frente a la $H_1 : \text{rango } A(1) = r$.

$$-T \ln(1 - \hat{\lambda}_r) \quad (7.2.9)$$

En este caso se contrasta si hay $r - 1$ relaciones de cointegración frente a la alternativa de que hay r , y por tanto el contraste es similar al desarrollado por Stock y Watson (1988).

Este área de la literatura está evolucionando muy rápidamente. El lector interesado puede encontrar comentados otros procedimientos de estimación y contrastación de hipótesis en Dolado *et. al.* ... (1989), Engle y Yoo (1989), Gonzalo (1989) y Phillips y Ouliaris (1988).

8. Extensiones

En este artículo nos hemos centrado en los resultados fundamentales de este área. Sin embargo, hemos dejado de lado muchos elementos y extensiones de interés. Entre ellos podríamos citar la aplicación de las técnicas de cointegración, y de los modelos de corrección de error, a *datos estacionales*. En el artículo de Hylleberg *et. al.* ... (1990) se desarrollan estos procedimientos con todo detalle. Los interesados en aplicar los contrastes de raíces unitarias a datos estacionales pueden consultar también Ilmakunnas (1990).

Otro área de interés trata de la generalización del concepto de cointegración a *modelos no-lineales*. Granger (1987) formaliza esta idea mediante los conjuntos de atracción. Escribano (1987,b) estudia las propiedades de funciones no lineales de variables integradas y aplica los resultados obtenidos al estudio de los *modelos de corrección de error no lineales*. En estos modelos la corrección de error la relación de cointegración es lineal y sin embargo es el ajuste hacia el equilibrio el que se permite que sea no lineal. Estos modelos han resultado ser

muy útiles para representar el comportamiento de la demanda de dinero en el Reino Unido desde 1878 hasta 1975, véase Escribano (1986), Escribano (1986, b) y Hendry y Ericsson (1989).

Otro área que necesita mayor desarrollo es el estudio de variables integradas cuyo orden de integración es mayor que uno. En este contexto las relaciones de cointegración polinomiales deberían jugar un papel relevante.

El punto más débil de la literatura de cointegración y raíces unitarias es la falta de poder de los contrastes de hipótesis. El desarrollo de técnicas más fiables debería ser el centro de atención en el futuro próximo.

Referencias

- ANDRES, J.; ESCRIBANO, A.; MOLINAS, C., y TAGUAS, D. (1990): «Modelización Econométrica con Restricciones de Equilibrio: La Inversión en España». Aceptado para su publicación en Antoni Bosch, editor.
- ANDREWS, D. W. K. (1988): «Least Squares Regression with Integrated or Dynamic Regressors under Weak Error Assumptions», *Econometric Theory*.
- BANERJEE, A.; DOLADO, J.; HENDRY, D. F., y SMITH, G. (1986): «Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence», *Oxford Bulletin of Economics and Statistics*, 48, 253-279.
- BEVERIDGE, S., y NELSON, C. R. (1981): «A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle», *Journal of Monetary Economics*, 7, 151-174.
- BEWLEY, R. (1979): «The Direct Estimation of the Equilibrium Response in a Linear Dynamic Model», *Economic Letters*, 3, 357-362.
- BILLINGSLEY, P. (1986): «Convergence of Probability Measures», John Wiley, Nueva York.
- BOX, G. P. E., y JENKINS, G. M. (1976): «Time Series Analysis, Forecasting and Control», Holden Day, San Francisco.
- COCHRANE, J. (1988): «How Big is the Random Walk in GNP?», *Journal of Political Economy*, 96, 893-920.
- DICKEY, D. A., y FULLER, W. A. (1979): «Distribution of Estimators for Autoregressive Time Series with a Unit Root», *Journal of American Statistical Association*, 74, 427-431.
- DICKEY, D. A., y FULLER, W. A. (1981): «Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root», *Econometrica*, 49, 1057-1072.
- DOLADO, J. (1989,a): «Optimal Instrumental Variable Estimator of the AR Parameter of an ARMA (1,1)». Aparecerá en *Econometric Theory*.
- , 1989,b: «Cointegración: Una Panorámica». Aparecerá en *Estadística Española*.
- DOLADO, J., y JENKINSON, T. (1987): «Cointegration: A Survey of Recent Developments». Documento de Trabajo del Banco de España.
- DOLADO, J.; JENKINSON, T., y SOSVILLA-RIVERO, S. (1989): «Cointegration and Unit Roots: A Survey». Documento de Trabajo del Banco de España, diciembre.
- DOLADO, J.; ANDRES, J., y DOMENECH, R. (1990): «Inferencia en Modelos Dinámicos Uniecuacionales con Variables Integradas», *Cuadernos Económicos de ICE*, N.º 44, 1990/1.
- ENGLE, R. F., y BOLLERSLEV, T. (1986): «Modelling the Persistence of Conditional Variances», *Econometric Reviews*, 5 (1), 1-50.
- ENGLE, R. F., y GRANGER, C. W. J. (1990): «Cointegración y Corrección de Error:

- Representación, Estimación y Contrastes», *Cuadernos Económicos de ICE*, Número 44, 1990/1.
- ENGLE, R. F., y YOO, B. S. (1987): «Forecasting and Testing in Cointegrated Systems», *Journal of Econometrics*, 35, 143-159.
- , 1989: «Cointegrated Economic Time Series: A Survey with New Results», *Discussion Paper*, University of California, San Diego, agosto, 1989.
- ESCRIBANO, A. (1986): «Identification and Modelling of Economic Relationships in a Growing Economy», *Ph. D. Dissertation*, University of California, San Diego.
- , 1986,b: «NonLinear Error-Correction: The Case of Money Demand in the U.K. (1878-1970)», Mimeo, Universidad de California, San Diego.
- , 1987,a: «Co-integration, Time Co-trends and Error Correction Systems: An Alternative Approach», *C.O.R.E., Discussion Paper*, 8715, Université Catholique de Louvain.
- , 1987,b: «Error-Correction Systems: Nonlinear Adjustments to Linear Long Run Relationships», *C.O.R.E., Discussion Paper*, 8730, Université Catholique de Louvain.
- , 1990: «Integration and Cointegration Under Structural Changes», Mimeo, Universidad Complutense de Madrid.
- FERNANDEZ MACHO, F. J.; HARVEY, A. C., y STOCK, J. H. (1987): «Forecasting and Interpolation Using Vector Autorregressions with Common Trends», *Annales D'Economie et de Statistique*, 6-7, 279-287.
- FRIEDMAN, M. (1973): «Una Teoría de la Función de Consumo», Alianza Universidad.
- FULLER, W. (1976): «Introduction to Statistical Time Series», John Wiley and Sons, Nueva York.
- GRANGER, C. W. J. (1981): «Some Properties of Time Series Data and Their Use in Econometric Model Specification», *Journal of Econometrics*, 16 (11), 121-130.
- , 1986: «Developments in the Study of Cointegrated Economic Variables», *Oxford Bulletin of Economic and Statistics*, 48, 3.
- GRANGER, C. W. J., y LEE, T. W. (1989): «Multicointegration». Aparecerá en *Advances in Econometrics*, 1989, eds.: G. F. Rhodes (Jr.), T. B. Fomby, Jai Press.
- GONZALO, J. (1989): «Comparison of Five Alternatives Methods of Estimating Long Run Equilibrium Relationships», *Discussion Paper*, 55, diciembre, University of California, San Diego.
- HALL, A. (1989): «Testing for a Unit Root in the Presence of Moving Average Errors», *Biometrika*, 76, 49-56.
- HENDRY, D. F., y ERICSSON, N. R. (1989): «An Econometric Analysis of UK Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz». Aparecerá en *American Economic Review*.
- HYLLEBERG, S.; ENGLE, R. F.; GRANGER, C. W. J., y YOO, B. S. (1990): «Integración y Cointegración Estacional», *Cuadernos Económicos de ICE*, N.º 44, 1990/1.
- JOHANSEN, S. (1988): «Statistical Analysis of Cointegrating Vectors», *Journal of Economic Dynamics and Control*, vol. 12, 231-254.
- JOHANSEN, S., y JUSELIUS, K. (1989): «Maximum Likelihood Estimation and Inference on Cointegration-with Applications to the Demand for Money». Aparecerá en *Oxford Bulletin of Economics and Statistics*.
- KAILATH, T. (1980): «Linear Systems», Prentice Hall.
- KING, R.; PLOSSER, C. I.; STOCK, J. H., y WATSON, M. W. (1987): «Stochastic Trends and Economic Fluctuations», *Working Paper*, 2229, abril, National Bureau of Economic Research.
- MARAVALL, A. (1987): «Descomposición de Series Temporales: Especificación, Estimación e Inferencia», *Estadística Española*, 114, 11-69.
- MOLINAS, C. (1986): «A Note on Spurious Regressions with Integrated Moving Average Errors», *Oxford Bulletin of Economics and Statistics*, 48.

- NEWKEY, W., y WEST, K. (1987): «A Simple, Positive Semi-definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix», *Econometrica*, 55, 703-708.
- PANTULA, S. G. (1989): «Testing for Unit Roots in Time Series Data», *Econometric Theory*, 5, 256-271.
- PEÑA, D. (1990): «Cointegración y Reducción de Dimensionalidad en Series Temporales Multivariantes», *Cuadernos Económicos de ICE*, N.º 44, 1990/1.
- PERRON, P. (1986): «Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach», *Discussion Paper*, University of Montreal.
- , 1989: «The Great Crash, the Oil Price Shock and the Unit Root Hypothesis», *Econometrica*, 57, 6, noviembre, 1361-1401.
- PHILLIPS, P. C. B. (1987): «Time Series Regression with a Unit Root», *Econometrica*, 55, 2, March.
- PHILLIPS, P. C. B., y OULIARIS, S. (1988): «Testi. ́ for Cointegration using Principal Components Methods», *Journal of Economic Dynamics and Control*, 12, 205-230.
- PHILLIPS, P. C. B., PERRON, P. (1988): «Testing for a Unit Root in Time Series», *Biometrika*, 75, 335-346.
- ROSENBLATT, M. (1978): «Dependence and Asymptotic Independence for Random Processes», *In Studies in Probability Theory*, M. Rosenblatt (ed.), Washington D. C., Mathematical Association of America.
- SAID, S., y DICKEY, D. (1984): «Testing for Unit Roots in Autorregressive-moving Average Models of Unknown Order», *Biometrika*, 71, 599-607.
- , 1985: «Hypothesis Testing in ARIMA (p, d, q) Models», *Journal of American Statistical Association*, 80, 369-374.
- SHILLER, R. J., y PERRON, P. (1985): «Testing the Random Walk Hypothesis: Power vs. Frequency of Observation», *Economic Letters*, 18, 381-386.
- SCHWERT, G. W. (1987): «Effects of Model Specification on Test for Unit Roots in Macroeconomic Data», *Journal of Monetary Economics*, 20, 73-103.
- STOCK, J. H. (1987): «Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors», *Econometrica*, 55, 381-386.
- STOCK, J. H., y WATSON, M. W. (1987): «Variable Trends in Economic Time Series», *Journal of Economic Perspectives*, 2, 147-174.
- , 1988: «Testing for Common Trends», *Journal of American Statistical Association*, 83, 1094-1107.
- WATSON, M. W. (1986): «Univariate Detrending Methods with Stochastic Trends», *Journal of Monetary Economics*, 18, 49-75.
- WEST, K. (1988): «Asymptotic Normality when Regressors have a Unit Root», *Econometrica*, 56, 1397-1411.
- WHITE, H. (1984): «Asymptotic Theory for Econometricians. In the Series on Economic Theory», *Econometrics and Mathematical Economics*, Academic Press, Nueva York.
- YOO, B. S. (1987): «Cointegrated Time Series: Structure, Forecasting and Testing», *Ph. D. Dissertation*, University of California, San Diego.